# Implicit Preference Labels for Learning Highly Selective Personalized Rankers

Paul N. Bennett
Microsoft
pauben@microsoft.com

Milad Shokouhi
Microsoft
milads@microsoft.com

Rich Caruana
Microsoft
rcaruana@microsoft.com

## ABSTRACT

Interaction data such as clicks and dwells provide valuable signals for learning and evaluating personalized models. However, while models of personalization typically distinguish between clicked and non-clicked results, no preference distinctions within the non-clicked results are made and all are treated as equally non-relevant.

In this paper, we demonstrate that failing to enforce a prior on preferences among non-clicked results leads to learning models that often personalize with no measurable gain at the *risk* that the personalized ranking is worse than the non-personalized ranking. To address this, we develop an implicit preference-based framework that enables learning highly selective rankers that yield large reductions in risk such as the percentage of queries personalized. We demonstrate theoretically how our framework can be derived from a small number of basic axioms that give rise to well-founded target rankings which combine a weight on prior preferences with the implicit preferences inferred from behavioral data.

Additionally, we conduct an empirical analysis to demonstrate that models learned with this approach yield comparable gains on click-based performance measures to standard methods with far fewer queries personalized. On three real-world commercial search engine logs, the method leads to substantial reductions in the number of queries re-ranked (2×-7× fewer queries re-ranked) while maintaining 85-95% of the total gain achieved by the standard approach.

## Categories and Subject Descriptors

H.3.3 [**Information Retrieval**]: Retrieval Models

## Keywords

Personalization risk, robust algorithms, re-ranking

## 1. INTRODUCTION

Personalizing search results based on context has been consistently reported to improve retrieval effectiveness [3, 27, 31, 34, 35]. However, personalization cannot help all queries and knowing when to *selectively* apply personalization is one of the key challenges of personalization [33]. In particular, when personalization is not necessary, personalizing the ranking runs the *risk* of decreasing performance relative to the non-personalized ranker. In order to personalize appropriately, we need an indication of the user's personal preferences as a target for learning.

Ideally one would obtain explicit judgments from each person for each personalized query, but that is not feasible at scale. A common alternative is to use "satisfied"[1] or long-dwell clicks to infer an implicit relevance judgment. In particular, the literature suggests that clicks indicate a relative preference over non-clicks but should not be interpreted as absolute relevance [1, 2, 22, 24]. As a result, a number of personalization studies have been conducted where the goal is to see a relative change in how high clicked results are ranked in the personalized versus non-personalized rankings [3, 4]. For example, an increase in the mean average precision of satisfied-clicked results (relevant) vs. the remaining results (non-relevant) over the non-personalized baseline indicates that, on average, the personalized ranker lists results users prefer higher in the rankings.

While technically correct, the lack of a prior on the many *unclicked* documents leads to models that often re-rank even when there are no demonstrable gains. This leads to a risk of personalization failure [37] not captured by click-based measures of risk. Furthermore, spuriously re-ranking when there is no need increases variance in the rankings. This variance masks the signal of performance improvements when any new improvement is tested – increasing the cost of interleaving [25] and A/B testing by requiring longer experiments and slowing development cycles. Ultimately, this occurs because click-based behavioral measures of relevance treat any ranking of the unclicked results as equivalent. We alleviate this by introducing a method which uses the non-personalized ranking to inform the target ranking of unclicked results in the absence of other information.

Table 1 presents an illustrative example of the problem of interpreting interactions as implicit relevance judgments. Here the user has been presented with a set of search results in response to the query `[acl]` in the ranked order of the first column (*Rank*) and clicked and dwelled on the sixth result for the "Association for Computational Linguistics" homepage. The table presents two hypothetical rankings, *A* and *B*, which both place the satisfied clicked item first but differ greatly in how the non-clicked results are ranked.

Ranking *A* leaves all of the remaining results in their original order; for both learning and evaluation this is highly conservative—if personalization is not appropriate, a set of users for whom the original non-personalized order was appropriate would find a desired result at most one position lower. Likewise, the variance from the original ranking or across people who may experience different personalized rankings is minimal at this conservative point. In contrast, after the satisfied clicked result is placed first, ranking *B* then inverts the order of the original first five results and then gives the last four results in their original order. Intuitively, if we believe the original

[1]We use the common definition of a "satisfied" click as a click with dwell of $\geq$ 30s or one that terminates the search session [7, 17].


*ICTIR'15,* September 27–30, Northampton, MA, USA.
 ISBN 978-1-4503-3833-2/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2808194.2809464.

**Table 1: Given ranked documents presented to a user with a click interaction given in "Satisfied Click", we show two possible re-rankings which order the documents as: A = $[6, 1, 2, 3, 4, 5, 7, 8, 9, 10]$ and B = $[6, 5, 4, 3, 2, 1, 7, 8, 9, 10]$. Both rankings have the same average precision (AP) when treating clicks as relevance judgments, but A is much more conservative in reordering the non-personalized ranking while B is much riskier. The columns on the right show two sets of gains derived according to the method in Section 2.1 that would both give rise to ranking A as the ideal ranking, but the column on the left places more weight on the original non-personalized ranking when computing the gains. See Section 2.1 for details.**

| Rank | Title and URL | Satisfied Click | Proposed Re-rankings | | Target Gain ($\alpha$,$\beta$) | |
|---|---|---|---|---|---|---|
| | | | A | B | (1,0.5) | (1,0.05) |
| 1 | Anterior cruciate ligament - Wikipedia, the free encyclopedia<br>en.wikipedia.org/wiki/Anterior_cruciate_ligament | No | 2' | 6' | 4.0 | 0.40 |
| 2 | Austin City Limits Music Festival - Official Site<br>www.aclfestival.com | No | 3' | 5' | 3.5 | 0.35 |
| 3 | ACL compliance, audit, governance & risk software<br>www.acl.com | No | 4' | 4' | 3.0 | 0.30 |
| 4 | Anterior Cruciate Ligament (ACL) Injuries-Topic Overview<br>www.webmd.com/a-to-z-guides/anterior-cruciate-ligament-acl-injuries-topic-overview | No | 5' | 3' | 2.5 | 0.25 |
| 5 | Access control list - Wikipedia, the free encyclopedia<br>en.wikipedia.org/wiki/Access_control_list | No | 6' | 2' | 2.0 | 0.20 |
| 6 | Association for Computational Linguistics \| ACL Homepage<br>www.aclweb.org/ | Yes | 1' | 1' | 9.0 | 9.0 |
| 7 | Association of Christian Librarians: Welcome<br>www.acl.org/ | No | 7' | 7' | 1.5 | 0.15 |
| 8 | About ACL - Administration for Community Living<br>www.acl.gov/About_ACL/Index.aspx | No | 8' | 8' | 1.0 | 0.10 |
| 9 | ACL Cargo<br>www.aclcargo.com | No | 9' | 9' | 0.50 | 0.05 |
| 10 | ACL Live, Austin, Texas<br>acl-live.com | No | 10' | 10' | 0 | 0 |

ranking is tuned for the overall population, ranking *B* is riskier since a high penalty is paid when the original non-personalized ranking is the correct intent. Even when correct, it introduces variance that can lengthen experiment time to determine statistical significance. Click-based measures of risk cannot capture distinctions among the order of the unclicked items, but other measures can – such as the percentage of queries re-ranked (personalized) and the correlation of the personalized rankings with the non-personalized rankings.

We seek to formally capture these intuitions about how to combine interaction data with the original ordering. We continue in the next section by establishing two simple axioms for deriving preference strengths from clicks. We demonstrate that when preference is defined according to these axioms, the pairwise preferences give rise to a target ranking with desirable properties.

# 2. PROBLEM APPROACH

We propose our model – referred to as *Weight-Initial-Pref* hereafter – based on a set of axioms which determine how we assign the strength of preferences between results based on their original presented position[2] and click-derived relevance. We demonstrate that when preferences are assigned in accordance with these axioms, the target ranking of results that can be derived from the preferences is constrained in terms of how far it can deviate from the non-personalized ranking. We then respect these constraints during training personalized rankers by using the target ranking derived from the preferences to learn a more conservative model.

In order to balance the search engine's non-personalized ranking and user interaction, we take a simple approach which encodes the strength of preferences between two search results as a function of the ranking and interaction. These pairwise strengths are then accumulated to each result to indicate the overall utility or gain that the result has. When sorted from greatest to least gain this yields a desired or *target* ranking with associated gains to use in learning a ranking. We demonstrate how starting from basic principles in designing the pairwise preference function, the final target ranking has several desirable properties.

## 2.1 Axioms for Stable, Personalized Ranking

As illustrated in Table 1, using whether a result received a satisfied click as a relevance label[3] does not distinguish between results in the same relevance class. It is this lack of a default order that ultimately gives rise to the variance in rankings when learning from clicks. An intuitive order to use as a default prior is the ordering given by the non-personalized ranking. We thus desire a way to incorporate this default order with the relevance signal from interaction and a way to increase or decrease the weight on the prior. In this section, we demonstrate a simple approach to achieving these goals.

More formally, we assume we have a set of results $\mathcal{D}$ that have been returned to the user in response to a particular query. Further, we assume an initial complete ordering $\Pi$ over the results; that is, $\Pi$ yields a consistent, transitive set of pairwise orderings of any two results $d_i, d_j \in \mathcal{D}$ which we indicate by $d_i \succ_\pi d_j$, to signify $d_i$ is preferred to $d_j$. In our setting $\Pi$ is the ranking of results that a user was presented with and $d_i$ being ranked "above" or "higher" than $d_j$ is indicated by $d_i \succ_\pi d_j$.[4]

We assume a setting where given the presented ranking $\Pi$ and a set of interactions, we would like to define a function, $\text{pref}(d_i \succ d_j)$, that indicates the strength of updated beliefs given the interactions consistent with the following two axioms:

[2]We typically refer to the non-personalized ranking as determining presentation order but the model easily applies to the case where the user interacted with a personalized ranker, and these interactions will be used to learn a new, updated personalized model.

[3]When we reach the empirical evaluation, those results with a satisfied click will be deemed to be in the relevant class while the remaining will be deemed to be non-relevant.

[4]Note that if $i$ and $j$ correspond to the ranks of the results in $\Pi$ then $d_i \succ_\pi d_j$ if and only if $i < j$.

1. The strength of preference for a relevant result over a non-relevant result should be stronger than any other preference.
2. For any labeled result from the same relevance class (relevant, non-relevant), the preference should reflect the preference of the presented ranking, $\Pi$.

The first of these axioms is commonly accepted in the literature. However, the second axiom is novel and essentially introduces the notion that: absent of any deciding behavioral signal from the user, the default preference should conservatively break ties by preferring the non-personalized ranking – which has benefited from being optimized over a large set of non-personalized relevance judgments.

In this paper, we assume the interactions partition the results into two sets, the satisfied clicks or relevant results, $\mathcal{R}$, and the non-satisfied clicks or non-relevant results, $\mathcal{I}$. We assume two user-defined parameters $\alpha$, $\beta$ such that $\alpha, \beta > 0$ where $\alpha$ indicates the preference for a relevant result over an irrelevant result and $\beta$ indicates the preference for maintaining the prior ranking for results in the same relevance class. We break the definition into two parts:

When $d_i \in \mathcal{R}$:

$$\text{pref}(d_i \succ d_j) = \begin{cases} \beta & d_j \in \mathcal{R}, d_i \succ_\pi d_j \\ 0 & d_j \in \mathcal{R}, d_i \prec_\pi d_j \\ \alpha & d_j \in \mathcal{I} \end{cases}$$

When $d_i \in \mathcal{I}$:

$$\text{pref}(d_i \succ d_j) = \begin{cases} 0 & d_j \in \mathcal{R} \\ \beta & d_j \in \mathcal{I}, d_i \succ_\pi d_j \\ 0 & d_j \in \mathcal{I}, d_i \prec_\pi d_j \end{cases}$$

The gain for a result $d_i$ is then defined to be:

$$G(d_i) = \sum_{d_j \in \mathcal{D}, d_i \neq d_j} \text{pref}(d_i \succ d_j) \tag{1}$$

where a higher gain is considered to be more highly relevant. That is, the gain for a result is simply the sum of the strength of preferences across all pairs of documents in the result set for this query.

In our approach we will maximize the normalized discounted cumulative gain (NDCG) [20] using these gains. We choose to do this since NDCG encourages placing the results with the highest gains high in the rankings, but one could also use the gains to optimize a measure that does not weight according to position in the ranking if desired. The ranking derived from sorting the results according to these gains can be considered to be the target ranking as far as a learning algorithm is considered.

*Properties of the Target Ranking.* To provide guidance in setting $\alpha, \beta$, we now consider what properties the target rankings *that are used for learning* have when the parameters take values satisfying $\alpha > \beta > 0$. This is important for demonstrating that the goal of optimization is sensible and meets our overall goals of improving personalized relevance while being conservative. We sketch the proofs of these statements briefly below.

First, we can prove that in the target ranking all results in the relevant set are above any from the non-relevant set (see Theorem 1 and Corollary 1). Additionally, we can prove that the target ranking is conservative in that when there are multiple results in the relevant set, the original non-personalized ranking is preserved among the relevant set in the target ranking (see Theorem 2 and Corollary 2). Finally, we can prove that the target ranking is conservative in that when there are multiple results in the non-relevant set, the original non-personalized ranking is preserved among the non-relevant set in the target ranking (see Theorem 3 and Corollary 3).

If $\beta > \alpha$ is allowed, target rankings can result where an irrelevant result has a higher gain than a relevant result. As $\beta$ increases, these violations of the first axiom occur more frequently in target rankings in the training set. We omit this from the experimental section but performance degrades almost immediately when $\beta > \alpha$, demonstrating the desirability of the first axiom.

*Proof Sketches.* We remind the reader of the assumption of an initial complete ordering $\Pi$ over the documents whose ordering of two documents $d_i, d_j \in \mathcal{D}$ is indicated by, $d_i \succ_\pi d_j$, to signify $d_i$ is preferred to $d_j$. In our setting $\Pi$ is the non-personalized ranking that a user was presented with and $d_i$ being ranked "above" or "higher" than $d_j$ indicates $d_i \succ_\pi d_j$.

THM. 1. *If $\alpha > \beta > 0$, $d_i \in \mathcal{I}$, and $d_j \in \mathcal{R}$ then $G(d_i) < G(d_j)$.*

$$\begin{aligned} G(d_i) &= \text{pref}(d_i \succ d_j) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_i \succ d_k) && (2) \\ &\quad \text{Since } d_i \in \mathcal{I}, d_j \in \mathcal{R}, \text{ and } \alpha > \beta > 0, \text{ then} \\ &\quad \forall d_k \, \text{pref}(d_i \succ d_k) \leq \text{pref}(d_j \succ d_k), \text{ and} \\ &\leq \text{pref}(d_i \succ d_j) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_j \succ d_k) && (3) \\ &\quad \text{Since } d_i \in \mathcal{I} \text{ and } d_j \in \mathcal{R}, \text{ then we have} \\ &\quad \text{pref}(d_i \succ d_j) < \text{pref}(d_j \succ d_i), \text{ yielding} \\ &< \text{pref}(d_j \succ d_i) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_j \succ d_k) && (4) \\ &= G(d_j). \end{aligned}$$

□

COROLLARY 1. *In the target ranking, all relevant documents are above all non-relevant documents. Proof: Follows trivially from Theorem 1 and sorting in descending order by gain.* □

THM. 2. *If $\alpha, \beta > 0$, $d_i \in \mathcal{R}$, $d_j \in \mathcal{R}$, and $d_i \succ_\pi d_j$, then $G(d_i) > G(d_j)$.*

$$\begin{aligned} G(d_i) &= \text{pref}(d_i \succ d_j) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_i \succ d_k) && (5) \\ &= \text{pref}(d_i \succ d_j) + \sum_{d_k \in \mathcal{R}, d_k \neq d_i, d_j} \text{pref}(d_i \succ d_k) \\ &\quad + \sum_{d_k \in \mathcal{I}, d_k \neq d_i, d_j} \text{pref}(d_i \succ d_k) && (6) \\ &\quad \text{Note we have } \forall d_k \in \mathcal{R} \text{ s.t. } d_k \succ_\pi d_j, \\ &\quad \text{pref}(d_j \succ d_k) = 0 \text{ and } \forall d_k \in \mathcal{R} \text{ s.t. } d_j \succ_\pi d_k, \\ &\quad d_i \succ_\pi d_k \text{ since } d_i \succ_\pi d_j \text{ and } \Pi \text{ is a complete ranking.} \\ &\quad \text{Therefore } \forall d_k, \text{pref}(d_i \succ d_k) \geq \text{pref}(d_j \succ d_k) \text{ yielding} \\ &\geq \text{pref}(d_i \succ d_j) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_j \succ d_k) && (7) \\ &\quad \text{Finally since } d_i, d_j \in \mathcal{R} \text{ and } d_i \succ_\pi d_j, \\ &\quad \text{pref}(d_i \succ d_j) > \text{pref}(d_j \succ d_i) \text{ yielding} \\ &> \text{pref}(d_j \succ d_i) + \sum_{d_k \in \mathcal{D}, d_k \neq d_i, d_j} \text{pref}(d_j \succ d_k) && (8) \\ &= G(d_j). \end{aligned}$$

□

COROLLARY 2. *In the target ranking, all relevant documents have the same order as in the initial ranking, $\Pi$. Proof: Follows trivially from Theorem 2 and sorting in descending order by gain.* □

THM. 3. *If $\beta > 0$, $d_i \in \mathcal{I}$, $d_j \in \mathcal{I}$, and $d_i \succ_\pi d_j$, then $G(d_i) > G(d_j)$. (The proof of this theorem follows the same basic approach as Theorem 2 and is omitted.)*

COROLLARY 3. *In the target ranking, all non-relevant documents have the same order as in the initial ranking,* $\Pi$. *Proof: Follows trivially from Theorem 3 and sorting in descending order by gain.* □

*Implications of Choices for* $\alpha$ *and* $\beta$. To provide further guidance in how a user can control risk via the parameters, we consider the interpretation of the values of $\alpha$ and $\beta$. As $\beta$ increases relative to $\alpha$, an increasing amount of cumulative gain across all results comes from enforcing the preference for the non-personalized ranking. Therefore, increasing $\beta$ can be viewed as being increasingly risk averse since a higher percentage of the cumulative gain can be increasingly achieved by returning the non-personalized ranking. This is illustrated in Table 1 by the two gain columns – each of which would give rise to the target ranking *A* in the table. When the ratio of $\beta$ to $\alpha$ is $0.5 : 1 = 1 : 2$ the non-personalized ranking achieves 0.80 NDCG@10 relative to the optimal ranking *A*. When we decrease $\beta$ relative to $\alpha$ to $0.05 : 1 = 1 : 20$ the non-personalized ranking only achieves 0.20 NDCG@10 relative to the optimal. Thus, fixing $\alpha$ and varying $\beta$ from 0 to $\alpha$ is a smooth way of creating increasingly risk-averse models. In the empirical section, we demonstrate this holds empirically as well.

## 2.2 Extension to Unexamined Results

Next, we consider the results that were not examined by the user and therefore not clicked due to lack of examination rather than lack of relevance. That is, we may want to treat unexamined results differently than results examined and intentionally not clicked (skips). In particular we may desire to weaken the constraint of the second axiom and not require the order of the *unexamined* results to be maintained in the target ranking.

To this end, we adapt earlier work to our setting. In particular, Radlinski & Joachims [24] used implicit preferences to learn preference-based models and achieved the best results by treating a clicked result as more relevant than both the unclicked results above it as well as the immediately next result when unclicked. Their findings gave rise to the cascade-model [11] of user interaction where a user scans from top to bottom and the results below the lowest click are not examined. As Radlinski & Joachims noted, however, if unexamined results are simply omitted when learning a personalized model, a machine learning model overfits the data by learning to "flip the ranking" since a click is nearly always the lowest (or second to lowest) result in the ranking. To correct this, they introduce a positivity weight constraint strictly greater than zero on the weight of a feature derived from the non-personalized ranking. Increasing the weight constraint in their model corresponded to placing more emphasis on the non-personalized ranking. We generalize their approach to be usable in situations where the learning algorithm cannot easily deal with weight-based constraints by limiting the results for each query in the training set to only those results from one position beyond the lowest click and above. Then, using the weighting model of Section 2.1 with $\beta > 0$ is highly similar to Radlinski & Joachims' introduction of a constraint and increasing $\beta$ is like increasing their weight constraint. Note that because there are almost no irrelevant results low in the rankings (which boost the gain of irrelevant results above them), $\beta$ can slightly exceed $\alpha$ for a small amount before target rankings end up with an irrelevant result with higher gain than a relevant result. Because empirically a larger range is permissible, in the empirical section we allow a greater range of exploration of the $\beta$ parameter for this method.

# 3. EMPIRICAL METHODOLOGY

We compare our approach (Weight-Initial-Pref) against several baselines from the literature on three large-scale datasets from commercial search engine logs. As in other works that attempt to reduce the risk of a retrieval method, the goal here is not to further increase relevance but to reduce re-ranking percentage and risk and yield a better tradeoff [9, 10, 37].

We implement our approach of using both the interaction data together with the weighted initial preferences within the LambdaMART framework [6]. Specifically, we aggregate the preferences according to Eq. 1 to determine gains using the definition of the strength of preference function, "pref" defined in Section 2.1. We then simply optimize NDCG using this definition of gain. Because it is primarily the ratio of $\alpha$ to $\beta$ which controls the emphasis placed on the original ranking, we fix $\alpha = 1$ and vary $\beta \in \{0.01, 0.02, \ldots, 0.09, 0.1, 0.2, \ldots, 1\}$. We select a particular model by selecting the value of $\beta$ that yielded the largest area under the gain vs. re-ranking percentage curve over validation data.

## 3.1 Performance Measures

We examine two primary measures of effectiveness which have been used by many others, mean average precision (MAP) and normalized discounted cumulative gain (NDCG) [20] where satisfied clicks are relevant and the remaining are non-relevant. Although such an approach is imperfect, we use it because changing both the optimization and evaluation methods together could lead to bias that favors the method we introduce. We use MAP as a measure of effectiveness in our first two datasets that only distinguish two degrees of relevance (non-relevant, relevant). We use NDCG in the WSCD '14 dataset, because the dataset publishers set that as the standard for that dataset. The dataset publishers have defined relevance labels of irrelevant, relevant, and highly relevant according to how long the user dwelled on a document. We describe MAP and the measures derived from MAP in detail and omit the details of NDCG whose derived measures are analogous.

*MAP and* $\Delta$ *MAP.* To remind the reader, the average precision (AP) of a ranking for a particular query is the average of the precision@$k$ at each position $k$ where a relevant document is ranked. Because the average is taken only with respect to where relevant documents occur, it encourages ranking relevant documents higher in the ranking. AP consequently has a value of 1 when all relevant documents are ranked above non-relevant documents. Note that in the results section we scale all measures in the $[0, 1]$ interval to $[0, 100]$ for ease of readability. The mean average precision (MAP) is the mean of the average precision across all queries.

We may also consider the mean change in average precision for a system $S$ relative to the non-personalized ranker baseline (BASE):

$$\Delta\,\mathrm{MAP} = \frac{1}{N}\sum_{q}[\mathrm{AP}_S(q) - \mathrm{AP}_{\mathrm{BASE}}(q)] \tag{9}$$

where $N$ is the total number of queries in the test set. This change is equivalent to the MAP of a system with the baseline's performance removed – meaning it is positive when personalization increases performance relative to the non-personalized ranker and negative otherwise. We focus on $\Delta$ MAP as our primary gain measure since it follows the same trends as MAP and proprietary reasons prevent releasing absolute MAP numbers on one dataset.

*Combining Effectiveness and Reranking.* In addition to mean average precision and NDCG, we will also use other measures that provide a more accurate view of the tradeoff between relevance and risk. In particular, we examine gain per query re-ranked, mean correlation with the non-personalized ranking, and the mean correlation per query re-ranked.

Each re-ranked query implies a risk of performance degradation. To normalize for the number of re-rankings, we can divide by the total number of queries for which a system generated a re-ranking different than the non-personalized ranker, $R_S$. When penalizing the difference in MAP this way, we refer to it as $\Delta$ MAP/R.

*Risk, Reward, and Gain.* Wang et al. [37] took an approach to risk by separating the differences in performance from a baseline ranker (*e.g.*, Eq. 9) into the queries where performance is improved $Q_+$ and the queries where performance is decreased $Q_-$. The overall gain can then be rewritten as a difference of the reward, which for MAP is $\frac{1}{N}\sum_{q\in Q_+}[\mathrm{AP}_S(q) - \mathrm{AP}_{\mathrm{BASE}}(q)]$, and the risk, $\frac{1}{N}\sum_{q\in Q_-}[\mathrm{AP}_{\mathrm{BASE}}(q) - \mathrm{AP}_S(q)]$. When we use the term "risk" in the empirical section we mean this specific click-based measure of risk – not the broader notions of risk discussed earlier in the paper. As in their work, the maximal risk, reward, and gain are achieved by simply maximizing gain. For ease of interpretation, we normalize changes in risk and gain by dividing by performance of the model with maximal gain (which also has maximal risk). We give the result as a percentage. While this performance-based definition of risk has become more common in the last several years (*e.g.*, as one of the key measures for the TREC Web Track [10]), there is a difference between using it for personalization and using it for standard ad-hoc relevance. This is because, in personalization, there are pseudo-nonrelevant judgments, but in the TREC Web Track setting, there are actual relevance judgments. In the same work, Wang et al. also looked at several other measures that implied risk even when relevance judgments and clicks were not available – namely the percentage of queries re-ranked and the mean Kendall's $\tau$ ranking correlation with the non-personalized ranking. We display both of these as well as the mean Kendall's $\tau$ over just those queries that are re-ranked ("% Re-ranked", K-$\tau$, and K-$\tau|\mathrm{R}$, respectively), but we focus on the percentage of queries that are re-ranked as a better reflection of a more general notion of risk.

## 3.2 Baselines

*Standard Gain.* First, we compare to optimizing ranking performance as has been usually done (e.g. [3]) – treating all satisfied clicks as relevant and the rest as non-relevant. We do this using LambdaMART which is an application of the LambdaRank approach [5] to gradient-boosted decision trees. Gradient-boosted decision trees have been very successful in a number of information retrieval tasks (*e.g.*, the Yahoo! Learning To Rank challenge Track 1 where LambdaMART was a key component [6]). We set the key parameters for LambdaMART to default settings appropriate for learning problems with similar amounts of data: we set number of leaves = 10, minimum documents per leaf = 2000, number of trees = 500, learning rate = 0.25, and used the validation set for pruning. Since all of the models are integrated to work with LambdaMART, to aid comparability we do not change these parameter settings. We refer to this as the *Standard Gain* approach.

*Naïve.* We can reduce the gain, risk, and re-ranking of a personalized model in a naïve way that gives up an amount of gain proportional to the amount of risk/re-ranking reduced. Namely, take the personalized *Standard Gain* model and apply it on a per-query basis by flipping a biased coin with probability $p$ and use the personalized *Standard Gain* model when the coin comes up "heads" and the non-personalized ranker for "tails". By selecting an appropriate $p \in [0, 1]$ any point on the line between the *Standard Gain* and the non-personalized ranker can be attained.

*Lowest Click Plus One.* To see the impact on re-ranking when the likely unexamined results were not included to stabilize the ranking, we implemented the method of Section 2.2. Again this was integrated into LambdaMART using the aggregated gains of Eq. 1 and the strength of preference function of 2.1. We again fix $\alpha = 1$ and vary $\beta \in \{0.1, 0.2, \ldots, 2\}$. To show this method in the best light, we chose to perform model selection for this approach by maximal gain over the validation set.

*Risk-sensitive Optimization.* As mentioned throughout the paper, we also could use the risk-sensitive optimization introduced by Wang et al. [37]. In particular, they decompose the change of gain relative to a baseline into risk and reward and introduce a parameter $\alpha_{risk} \geq 0$ such that increasing $\alpha_{risk}$ makes the learning algorithm learn increasingly low risk models. We implement this in our setting by optimizing over the training set of queries, $Q$, the reward minus the weighted risk:

$$\mathrm{RS}(Q, \alpha_{risk}) = \frac{1}{N} \cdot [\sum_{q\in Q_+}[\mathrm{AP}_S(q) - \mathrm{AP}_{\mathrm{BASE}}(q)]] \quad (10)$$
$$-(\alpha_{risk}+1)[\sum_{q\in Q_-}[\mathrm{AP}_{\mathrm{BASE}}(q) - \mathrm{AP}_S(q)]].$$

Here $\alpha_{risk} = 0$ is equivalent to the *Standard Gain* model. Since Wang et al. demonstrated both a reduction in the number of queries re-ranked and an increase in correlation with the non-personalized ranking with increasing $\alpha_{risk}$ (at a cost to a less than proportional reduction in gain), this method makes a natural baseline. On the validation set we explore $\alpha_{risk} \in \{0, 0.1, 0.2, \ldots, 0.9, 1, 2, 3, \ldots 10\}$. We select a particular model by selecting the value of $\alpha_{risk}$ that yielded the largest area under the gain vs. re-ranking percentage curve over validation data. We also consider how re-ranking would compare if we compared a risk-sensitive model at the same gain achieved as the *Weight-Initial-Pref* model. We do this and present the results for this as *RS (Min Risk, Gain Parity)*. Likewise, we could use the risk-sensitive model that has the same re-ranking as the *Weight-Initial-Pref* model and observe the difference in gain. We present results for this as *RS (Max Gain, near Rerank)*.

# 4. DATA & FEATURES

We evaluate the methods on three datasets. The first is a proprietary dataset from Bing, Microsoft's search engine. The second and third are anonymized public datasets from the Yandex search engine first released to participants of the Relevance Prediction Challenge[5] which was part of the WSDM 2012 WSCD workshop and for the Personalized Web Search Challenge[6] organized in conjunction with the WSDM 2014 WSCD workshop. We refer to these testbeds, respectively, as the *WSCD '12* and *WSCD '14* datasets.

The Bing dataset consists of queries sampled between 25 October 2013 and 14 November 2013 (three weeks). We use the first week of data for training and the last two respectively for validation and testing. The training/validation/test sets contain 449K/444K/443K queries respectively. Because we have full access to this dataset we are able to compute a variety of short- and long-term features for personalization studied elsewhere in the literature. In particular, we implemented the features studied by others for long-term personal navigation in [34], location features in [4], and short- and long-term topical and navigation features as in [3]. Since our focus is not on features we refer the reader to those articles for details; it is only important to us that the *Standard Gain* model represent

[5] http://imat-relpred.yandex.ru/en

[6] https://www.kaggle.com/c/yandex-personalized-web-search-challenge

a competitive and realistic baseline for personalization from the literature. In particular, in order to have non-trivial relationships between gain and risk at the model level, the feature set should contain a rich set of features with the potential for the trade-off to be exploited.

To complement the analysis over the Bing data, the WSCD '12 and WSCD '14 datasets give two publicly available datasets to reproduce a similar style of experiment where the amount of personalization is limited by the types of features that are available. In particular in the WSCD '12 dataset only short-term (session) identifiers for users are available while the WSCD '14 dataset has both short- and long-term information via a consistent user ID across sessions. In both datasets, query text and URLs have been replaced with numerical IDs and topic and location information are not available. Thus, some well-studied personalization features in the literature dependent on location and topic cannot be computed. However, we can compute proxies for many studied features and have implemented the proxies for short-term refinding, personal navigation, and query similarity features described in [30] for the WSCD '12 dataset and both short- and long-term proxies of the same for the WSCD '14 dataset. We note that, on all datasets, we have the original position and score of documents under the ranking presented to the user as both a re-ranking feature and to compute our preference models.

For the WSCD '12 dataset, we split the data according to the SessionID metadata in the logs. Sampled sessions with SessionID smaller than 3E + 07 were used for training and validation, and the remainder were used for testing. In total, the sampled train/validation/test sets contain 593K/591K/592K queries respectively. We performed a similar partitioning by user ID over the published training set for WSCD '14 and downsampled to obtain train/validation/test sets of 1.31M/654K/654K queries, respectively.

## 5. RESULTS AND DISCUSSION

*Performance Summary.* We start by discussing results on the Bing dataset in Table 2 (top) . First, we note that all of the personalization models selected achieve significant gains in $\Delta$ MAP over the non-personalized baseline. Furthermore, the baseline *Standard Gain* model demonstrates similar performance to reported models in the literature that use both short- and long-term features [3]; However, it also re-ranks quite often (41.19%), and when the change in MAP is penalized by dividing by the number of queries re-ranked the credit drops to a difference of 2.58 ($\Delta$ MAP/R). This re-ranking can also be seen by Kendall's $\tau$ where the average overall (K-$\tau$) is lower than all of the other models except the *Lowest-Click-Plus-One* model even though the average Kendall's $\tau$ over only those queries re-ranked (K-$\tau|$R) for the *Standard Gain* model is comparable to those other models: this indicates when the model does change the ranking, the number of pairwise swaps relative to the non-personalized ranking is about the same. However, it changes the rankings far more often with no measurable gain. Both our method, *Weight-Initial-Pref* and the *Risk Sensitive* models improve the tradeoff of gain to percentage of queries re-ranked, but *Weight-Initial-Pref* is able to reduce the percentage of queries far more while maintaining gain. In the following discussion section, we describe more details about the differences in these methods.

When examining the *Lowest-Click-Plus-One* model, large gains relative to the non-personalized ranking are achieved, but a large number of queries are re-ranked (62.02%), and when changed, they are reordered on average the most of any method (K-$\tau|$R=86.68). As a result, a favorable tradeoff relative to the *Standard Gain* model is never attained. One possibility is that this model is simply increasing the variance among the unexamined documents; however, this exploration comes at a noticeable cost to measurable relevance.

Summarizing the results on WSCD '12 and WSCD '14 (respectively middle and bottom sections in Table 2), we note the same trends generally hold up although they are less pronounced in these datasets where less rich personalization features are available. Once again the personalized *Standard Gain* model produces gains over the non-personalized baseline ranking presented to the user. We suspect that the absence of many personalization signals (due to the anonymized nature of the release) is why none of the models reach as high of a level of effectiveness for low re-ranking as observed in the Bing data. Comparing between models, in the WSCD '12 dataset the *Weight-Initial-Pref* model again demonstrates a substantially higher re-ranking penalized change in MAP ($\Delta$ MAP/R $= 2.48$) while maintaining 84.60% of the gain. This is while reducing re-ranking by a 2× factor from the *Standard Gain* model of 98.14% and a 1.33× factor from the most comparable gain *Risk Sensitive* model of 65.35%. Results trend similarly on WSCD '14.

*Discussion.* The trade-offs between gains and percentage of queries reranked are illustrated in Figures 1 - 3 respectively for Bing, WSCD '12 and WSCD '14 datasets. In all these figures, the $x$-axis represents the percentage of re-ranked queries while the $y$-axis shows the performance of different methods in terms of *normalized gain*. The left plots in each figure are generated based on models *selected* for each approach while the right plots depict the gain vs. reranking trade-offs across all parameter values. Next we describe our selection criterion for picking these models.

Examining the left figures for *Naïve* baseline, as $p \rightarrow 1$, the resulting mixed model has a performance that moves to the top right and as $p \rightarrow 0$, the mixed model is increasingly the non-personalized ranking that has no relative improvements. The *Naïve* baseline represents the achievable performance by fixing $p$ to some value in $[0, 1]$. Therefore, only improvements above and to the left of the *Naïve* line represent useful tradeoffs not dominated by some *Naïve* model's performance with appropriately chosen $p$. More generally, for any two models, the tradeoffs along the line between them can be achieved in the same fashion. This means that increasing the convex hull[7] is better, and in particular, having a set of models whose area under the curve is maximized.

On the Bing dataset (Figure 1) we see that both the *Weight-Initial-Pref* models and *Risk Sensitive* models achieve improved tradeoffs of gain and percentage of queries re-ranked relative to the *Naïve* method. However, relative to all of the risk sensitive models, the *Weight-Initial-Pref* model offers substantial improvements, maintaining 95.28% of the total gain while only re-ranking 5.92% of the queries (See Table 2 for details). This is a 7× reduction in number of queries re-ranked relative to the *Standard Gain* model while keeping nearly all of the gain. As can be seen from the impact on queries where the SAT clicked results changed position, this was achieved by learning a model where the changed queries have a larger average gain than the *Standard Gain* model, and when penalized for re-ranking, the amount of improvement per query re-ranked ($\Delta$ MAP/R $= 17.08$) is much better than the other methods.

In comparing the *Weight-Initial-Pref* model to various risk sensitive tradeoffs, we see that if we try to choose a risk sensitive model with a similar gain but minimum risk, *RS (Min Risk, Gain Parity)*, we re-rank 3.5× more often (5.92% vs. 20.77%). This is the horizontal space between the black diamond and orange plus in Figure 1 (*left*). If we try to choose a risk sensitive model with a similar amount of re-ranking but maximal gain, *RS (Max Gain, near Rerank)*, we never reach as little re-ranking before gain starts dropping precipitously for 81.54% gain at 8.43% of re-ranking for the risk model vs.

[7]The outer edge not below a line connecting any other two points.

**Table 2: Results on Bing (*top*), WSCD '12 (*middle*), and WSCD '14 (*bottom*). There are 443k, 592K, and 654K queries in the test sets, respectively. Statistically significant ($p \leq 0.05$ two-tailed paired $t$-test) gains relative to the non-personalized ranker are underlined.**

| Method | Param. | Δ MAP | Δ MAP/R | % Re-ranked | % Risk | % Gain | K-$\tau$ | K-$\tau$\|R |
|---|---|---|---|---|---|---|---|---|
| Standard Gain | $\alpha_{risk}=0$ | 1.06 | 2.58 | 41.19 | 100.00 | 100.00 | 95.43 | 88.90 |
| Weight-Initial-Pref | $\beta=0.2$ | 1.01 | 17.08 | 5.92 | 59.27 | 95.28 | 99.39 | 89.70 |
| Risk Sensitive | $\alpha_{risk}=5$ | 0.95 | 9.46 | 10.04 | 37.87 | 89.39 | 99.02 | 90.23 |
| RS (Min Risk, Gain Parity) | $\alpha_{risk}=1.6$ | 1.01 | 4.87 | 20.77 | 57.95 | 95.21 | 97.83 | 89.55 |
| RS (Max Gain, near Rerank) | $\alpha_{risk}=10$ | 0.87 | 10.27 | 8.43 | 26.99 | 81.54 | 99.24 | 90.98 |
| Lowest-Click-Plus-One | $\beta=1.2$ | 0.93 | 1.50 | 62.02 | 50.46 | 87.53 | 91.74 | 86.68 |

| Method | Param. | MAP | Δ MAP | Δ MAP/R | % Reranked | % Risk | % Gain | K-$\tau$ | K-$\tau$\|R |
|---|---|---|---|---|---|---|---|---|---|
| Standard Gain | $\alpha_{risk}=0$ | 68.12 | 1.43 | 1.46 | 98.14 | 100.00 | 100.00 | 80.82 | 80.46 |
| Weight-Initial-Pref | $\beta=0.4$ | 67.90 | 1.21 | 2.48 | 49.01 | 55.34 | 84.60 | 96.18 | 92.21 |
| Risk Sensitive | $\alpha_{risk}=1.7$ | 67.90 | 1.22 | 1.87 | 65.35 | 43.37 | 85.19 | 92.60 | 88.68 |
| RS (Min Risk, Gain Parity) | $\alpha_{risk}=1.7$ | 67.90 | 1.22 | 1.87 | 65.35 | 43.37 | 85.19 | 92.60 | 88.68 |
| RS (Max Gain, near Rerank) | $\alpha_{risk}=3$ | 67.66 | 0.97 | 1.92 | 50.82 | 25.24 | 67.93 | 95.39 | 90.93 |
| Lowest-Click-Plus-One | $\beta=0.9$ | 67.02 | 0.34 | 0.34 | 99.87 | 28.15 | 23.72 | 74.87 | 74.84 |

| Method | Param. | NDCG@10 | Δ NDCG@10 | Δ NDCG@10/R | % Re-ranked | % Risk | % Gain | K-$\tau$ | K-$\tau$\|R |
|---|---|---|---|---|---|---|---|---|---|
| Standard Gain | $\alpha_{risk}=0$ | 80.54 | 0.8255 | 0.88 | 93.99 | 100.00 | 100.00 | 88.38 | 87.64 |
| Weight-Initial-Pref | $\beta=0.01$ | 80.45 | 0.7324 | 4.24 | 17.27 | 55.20 | 88.72 | 97.74 | 86.91 |
| Risk Sensitive | $\alpha_{risk}=3$ | 80.41 | 0.6949 | 4.21 | 16.50 | 32.37 | 84.18 | 98.07 | 88.30 |
| RS (Min Risk, Gain Parity) | $\alpha_{risk}=2$ | 80.45 | 0.7317 | 3.39 | 21.58 | 40.39 | 88.64 | 97.54 | 88.60 |
| RS (Max Gain, near Rerank) | $\alpha_{risk}=3$ | 80.41 | 0.6949 | 4.21 | 16.50 | 32.37 | 84.18 | 98.07 | 88.30 |
| Lowest-Click-Plus-One | $\beta=1$ | 80.13 | 0.4162 | 0.42 | 99.98 | 45.45 | 50.42 | 79.7 | 79.70 |

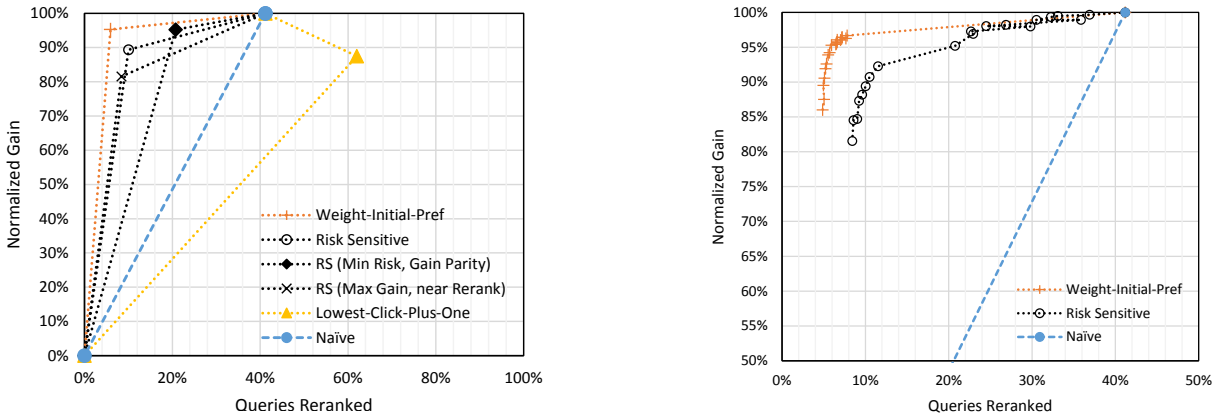


**Figure 1: In the Bing dataset, trade-off between gain (normalized to max of standard model) and queries re-ranked for selected models (*Left*) and at all parameter values for the Risk-Sensitive and Weight-Initial-Pref approaches (*Right*).**

95.28% gain at 5.92% re-ranking for the *Weight-Initial-Pref* model (vertical distance between the black 'x' and orange '+' in Figure 1). If we select the risk sensitive model to maximize the area for gain vs. re-ranked, both gain and re-ranking suffer; the risk model has 89.39% gain with 10.04% re-ranking vs. 95.28% gain at 5.92% re-ranking for the *Weight-Initial-Pref* model (diagonal distance between the black 'o' and orange '+' in Figure 1). In addition, while the *Weight-Initial-Pref* model is slightly worse in measurable risk than the *Risk Sensitive* model (59.27% to 57.95%) the comparable gain indicates that there has been an increase in measurable reward. Thus, for a slight increase in click-based risk there is a corresponding increase in click-based reward and a major reduction in re-ranking. Furthermore, we see in Figure 1 across the whole range of gain vs. re-ranking tradeoffs that the *Weight-Initial-Pref* models perform at least as well as the risk models and substantially outperform them in the low re-ranking, high gain corner (the optimal top left corner).

On the WSCD datasets, the *Weight-Initial-Pref* models also achieve better gain for comparable re-ranking across the full range in Figure 2 (*right*). In the WSCD '14, the impact is the lowest, here the edge of gain to re-ranking percentage is the least but still visible in Figure 3. In comparison to the risk-sensitive model that attains the same gain, *RS (Min Risk, Gain Parity)*, there is still a 20% reduction in re-ranking and a 5× reduction relative to the *Standard Gain* model. We believe improvements on this dataset could be increased by generalizing our method to preserve order between degrees of relevance – this is the only dataset with multiple degrees of relevance.

In summary, on all datasets the *Weight-Initial-Pref* models show large decreases in the amount of re-ranking relative to the baseline personalized *Standard Gain* models while reducing measurable gain slightly. It also provides moderate to large reductions in re-ranking relative to the risk-sensitive and *Lowest-Click-Plus-One* models while providing as much to substantially more gain than them.

## 6. RELATED WORK

Our work is related to prior research in personalization, learning to rank, axiomatic IR, click modeling, and online learning. We briefly review the most relevant work in each of these areas next.

*Search Personalization.* Personalizing search results based on context has been consistently reported to improve retrieval effective-

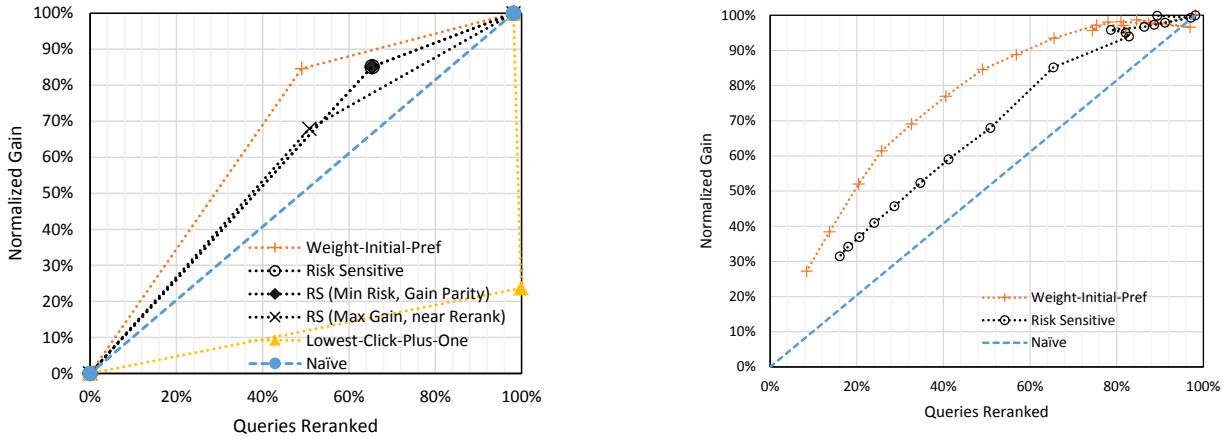


**Figure 2: In the WSCD '12 dataset, trade-off between gain (normalized to max of standard model) and queries re-ranked for selected models (*Left*) and at all parameter values for the Risk-Sensitive and Weight-Initial-Pref approaches (*Right*).**

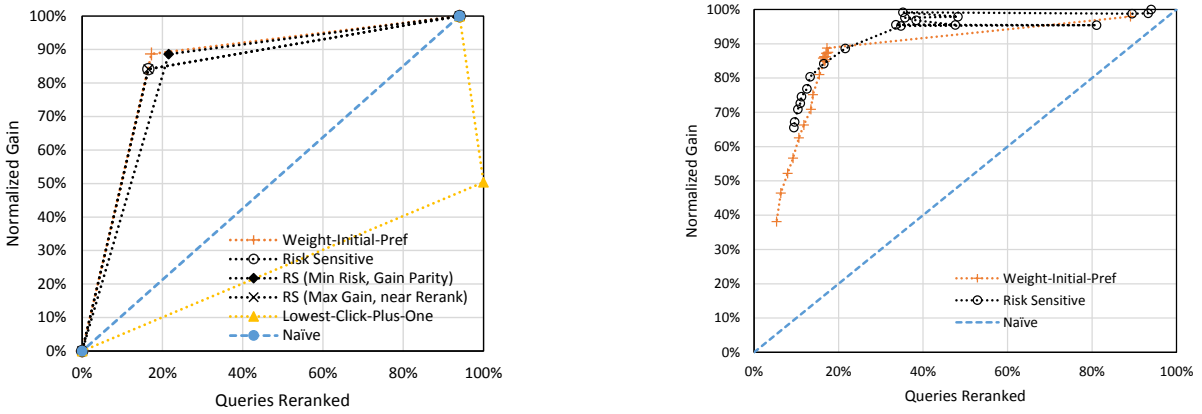


**Figure 3: In the WSCD '14 dataset, trade-off between gain (normalized to max of standard model) and queries re-ranked for selected models (*Left*) and at all parameter values for the Risk-Sensitive and Weight-Initial-Pref approaches (*Right*).**

ness [3, 27, 31, 34, 35]. Typically, *context* refers to anything that can distinguish the user from the rest – or most – of the population. The most widely used contextual features are users' previous queries and clicks [3, 30, 34, 39], but other contextual features based on a user's browsing history [35], user-specific topical profiles [3, 31], location [4] or demographics [29] have also been demonstrated to be effective for personalizing search results.

In the majority of these studies, the original results returned by a context-independent ranker are passed to a *re-ranker* trained for personalization. Like those works, we assume throughout that personalization will be conducted by learning a single model which re-ranks the results from a non-personalized ranker [3, 34, 35], but our technique is equally applicable for learning a single re-ranking model per user or constructing a target ranking for evaluation by other personalization frameworks. Such re-rankers are typically trained by sampling sessions from search logs and consider *satisfied* clicks in those sessions as ground-truth relevance.

What has often been overlooked by these techniques is the potential pairwise preferences between non-SAT or non-clicked documents. By treating all such documents as equally non-relevant, the personalized re-ranking models ignore the underlying pairwise preferences between such documents that could be inferred by respecting the positions in the original non-personalized ranking. In this paper, we demonstrate how integrating preferences based on the original ordering and click modeling can significantly reduce the risk of wrong re-rankings in personalization.

*Multi-objective Learning to Rank.* Learning to rank [23] covers a large body of supervised and semi-supervised techniques in which the goal is to *learn* a ranking function over available retrieval features. The labels for training are usually collected manually for a set of documents, or as stated earlier, are inferred based on collected clicks. The common objective among ranking models is to optimize for relevance. However, in many ranking scenarios, there might be more than one measure for optimization. For instance, Dong et al. [13] demoted outdated labels in their training data and developed a ranker that can be optimized for freshness and relevance. In a similar vein, Svore et al. [32] used historical clicks to break the ties in training when a pair of documents share the same relevance label.

Our work is related to these *multi-objective* scenarios, as we are interested in maximizing the personalization gain while minimizing the risk of diverging from the original preferences. The closest study to our work, which we also use as one of our experimental baselines, is the risk-sensitive optimization framework of Wang et al. [37] that is trained to maximize the difference between the total improvement in ranking relative to a baseline and the weighted total decrease in performance. Increasing the weight penalizes failures more heavily to learn more risk-averse models. Quite recently Dinçer et al. [12] extended this framework to a query-specific risk-weighting where the

weight was derived from the significance of deviation in risk relative to the overall risk distribution. In experimentation, we demonstrate that only optimizing a click-based measure of risk in personalization using the uniform weighting of queries as in [37] yields an inferior overall solution viewed according to several other measures of risk – such as percentage of queries re-ranked. Furthermore, although we do not pursue it here, our approach could be combined with either the optimization of the uniform query-weighting of risk or the query-specific weighting by using the implicit preferences to infer a target ranking and the optimization framework to optimize the risk-objective relative to the inferred target ranking.

*Axiomatic Information Retrieval.* Our work is related to other axiomatic approaches in information retrieval. For example, Fang and Zhai [15] proposed a set of constraints for weighting terms in documents based on document length and term statistics. The same authors later generalized term weighting models to incorporate the semantic similarity of terms [16]. Gollapudi and Sharma [18] discussed a set of axioms for balancing novelty and relevance in diversification and showed that no diversity function can satisfy them all. We develop an axiomatic approach for *personalization* where interaction feedback is combined with non-personalized relevance.

*Click Modeling.* Clicks capture a user's implicit feedback about documents and have been shown to provide effective ranking features [1] and large-scale pseudo-relevance labels [22]. Consequently, much attention has been devoted to click modeling and interpreting clicks in search logs. Joachims [21] pioneered the application of clicks as labels for optimizing rankers. In a follow up study, Joachims et al. [22] showed that clicks are subject to various presentation biases and proposed a set of rules for inferring pairwise preferences based on clicks. In another work, Agichtein et al. [2] reported similar biases and showed accounting for these biases can significantly improve retrieval effectiveness. Agichtein et al. [1] incorporated features based on click data in training rankers and proposed a set of features for predicting relevance preferences from clicks. We leverage these earlier insights to establish axioms for constraining the *strength* a preference should have; by comparing to an adaptation of the constrained preference ranking of Radlinski and Joachims [24], we demonstrate that which preferences are included for learning are key in controlling the risk of personalization.

Several studies have attempted to model user search behavior for better interpretation of clicks and separating out the relevance aspect from position bias [11, 14, 19]. In one of the earliest work in this area, Craswell et al. [11] suggested a *cascade* model in which users browse results from top to bottom and leave as soon as they find a document that satisfies their intent. The dynamic Bayesian network model (DBN) of Chapelle and Zhang [7] explains the clicks on documents based on their *perceived* and actual relevance. The former factor is determined according to the probability of click based on the URL, while the latter measures the probability of satisfaction given that the document is clicked. A recent work by Chuklin et al. [8] presents a comparative analysis over several state-of-the-art click models. When learning from clicks, Ustinovskiy and Serdyukov [35] investigate a simple fusion method of the scores of the personalized and non-personalized ranker to avoid over-personalizing. This is like our naïve baseline (see Section 3.2) although their method is targeted at increasing the correlation of the personalized ranking with the non-personalized ranking and not at optimizing the trade-off between the percentage of queries personalized and total gain. We demonstrate both increased correlation and a reduction in the percentage of queries personalized. Very recently, Ustinovskiy et al. [36] learned the weights to give to URLs during training by using interaction signals only available at training time. However, they do not demonstrate a reduction in risk and show a very marginal gain in relevance. Whether their method of using training time signals to learn weights can be incorporated with our approach is an interesting avenue for future work. In contrast to previous click modeling work, our focus is not only on predicting what will be clicked by the user, but we also enforce constraints during learning that ensure relevant documents are at the top while documents deemed non-relevant remain in their most conservative order.

*Online Learning.* Several recent approaches to online learning from interaction data are also related. In particular, Shivaswamy and Joachims [28] present and Raman et al. [26] later extend a coactive learning method for personalizing rankings from implicit feedback by using online learning to incrementally update a trained model from user click data. In their approach, the user feedback is in the form of clicks that are used to perform an online update to the model in order to improve the search results for the following searches. Perhaps the main limitation of the coactive learning method is the dependence on a linear weight vector over features; often complex non-linear models such as boosting [38] perform substantially better than linear models, and it is not clear how to extend the Coactive learning method to non-linear models that do not depend on linear weight vectors. In contrast, our method can be used by a variety of learning algorithms including gradient-boosted decision trees. Also, in order to make coactive learning robust to imperfect feedback, the rankings presented to users must be slightly perturbed to promote unbiased exploration [26]. In contrast, the method we propose in this paper avoids the problem of oscillation by learning personalized rankings that are consistent with the original pre-personalized ranking over all the data instead of in an online fashion — thus preventing the large changes in ranking that can occur with coactive learning in the presence of noisy user feedback.

In contrast to previous work, we demonstrate that click-based approaches to evaluating personalization do not capture all aspects of the risk of decreasing relevance by personalizing results when not appropriate. In particular, we focus on what axioms must be satisfied in determining the strengths of preferences to provide both a personalized signal of relevance while maintaining an overall conservative approach. We develop a framework that can be flexibly used to set gains for learning personalized models with a wide variety of learning methods including both linear models and gradient-boosted decision tree approaches. Empirically, we demonstrate that risk can be managed in personalization by incorporating information from the non-personalized ranking *while still maintaining* the constraint that clicked or satisfied clicked results are ranked highest.

## 7. FUTURE WORK & CONCLUSIONS

This work has many potential future extensions of interest. In particular, the strength of preference functions we give here are only one way to satisfy the axioms. There are other ways that would yield a weight on the non-personalized ranker preference ($\beta$) less than ($\alpha$). This includes but is not limited to making ($\beta$) query-specific where the number of past impressions might differentially weight updates to head or tail queries, dealing with query result churn by using the number of times a pair is displayed rather than the query, and using a Bayesian approach that models the strength as a probability and performs a Bayesian update depending on other factors (*e.g.*, eye and or cursor tracking). Likewise, considering how to integrate explicit preference judgments or multiple grades of relevance from either the user or an annotator are interesting extensions and might be used as the basis of considering other weightings on the unexamined results

after the lowest click. These extensions result directly from casting the problem as a preference assignment problem – while directly defining the target ranking given interaction feedback may be hard, it is both more principled and easier to define how to update pairwise preferences based on those interactions and how to aggregate the preferences to define a target ranking.

In summary, we identified a problem previously unappreciated in the literature. Namely, treating implicit data as relevance judgments leads to often re-ranking queries when there is no demonstrable gain. After identifying this problem, we formulated two simple axioms that lead to global constraints on a target ranking which has the relevant results always on top and then orders results within type consistent with the original ranking results after that. In three real-word commercial search engine logs, learning using this target ranking leads to a substantial reduction in the number of queries re-ranked of 2×-7× fewer queries re-ranked while maintaining 85-95% of the total gain achieved by the standard approach.

## References


[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. SIGIR*, 2006.

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. SIGIR*, 2006.

[3] P. Bennett et al. Modeling the impact of short- and long-term behavior on search personalization. In *Proc. SIGIR*, 2012.

[4] P. N. Bennett, F. Radlinski, R. W. White, and E. Yilmaz. Inferring and using location metadata to personalize web search. In *Proc. SIGIR*, 2011.

[5] C. Burges, R. Ragno, and Q. Le. Learning to rank with non-smooth cost functions. In *Proc. NIPS*, 2006.

[6] C. J. C. Burges, K. M. Svore, P. N. Bennett, A. Pastusiak, and Q. Wu. Learning to rank using an ensemble of lambda-gradient models. *JMLR*, 14, 2011.

[7] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proc. WWW*, 2009.

[8] A. Chuklin, P. Serdyukov, and M. de Rijke. Click model-based information retrieval metrics. In *Proc. SIGIR*, 2013.

[9] K. Collins-Thompson. Reducing the risk of query expansion via robust constrained optimization. In *Proc. CIKM*, 2009.

[10] K. Collins-Thompson, P. N. Bennett, F. Diaz, C. Clarke, and E. M. Voorhees. Trec 2013 web track overview. In *TREC '13*, 2013.

[11] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proc. WSDM*, 2008.

[12] B. T. Dinçer, C. Macdonald, and I. Ounis. Hypothesis testing for the risk-sensitive evaluation of retrieval systems. In *Proc. SIGIR*, 2014.

[13] A. Dong et al. Time is of the essence: Improving recency ranking using twitter data. In *Proc. WWW*, 2010.

[14] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proc. SIGIR*, 2008.

[15] H. Fang and C. Zhai. An exploration of axiomatic approaches to information retrieval. In *Proc. SIGIR*, 2005.

[16] H. Fang and C. Zhai. Semantic term matching in axiomatic approaches to information retrieval. In *Proc. SIGIR*, 2006.

[17] S. Fox et al. Evaluating implicit measures to improve web search. *ACM Trans. Inf. Syst.*, 23(2), Apr. 2005.

[18] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *Proc. WWW*, 2009.

[19] F. Guo, C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. Click chain model in web search. In *Proc. WWW*, 2009.

[20] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4), Oct. 2002.

[21] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. SIGKDD*, 2002.

[22] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.*, 25(2), Apr. 2007.

[23] T.-Y. Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3), 2009.

[24] F. Radlinski and T. Joachims. Query chains: Learning to rank from implicit feedback. In *Proc. SIGKDD*, 2005.

[25] F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality. In *CIKM '08*, 2008.

[26] K. Raman, T. Joachims, P. Shivaswamy, and T. Shnabel. Stable coactive learning via perturbation. In *Proc. ICML*, 2013.

[27] X. Shen, B. Tan, and C. Zhai. Context-sensitive information retrieval using implicit feedback. In *Proc. SIGIR*, 2005.

[28] P. Shivaswamy and T. Joachims. Online structured prediction via coactive learning. In *Proc. ICML*, 2012.

[29] M. Shokouhi. Learning to personalize query auto-completion. In *Proc. SIGIR*, 2013.

[30] M. Shokouhi, R. W. White, P. Bennett, and F. Radlinski. Fighting search engine amnesia: Reranking repeated results. In *Proc. SIGIR*, 2013.

[31] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *Proc. CIKM*, 2007.

[32] K. M. Svore, M. N. Volkovs, and C. J. Burges. Learning to rank with multiple objective functions. In *Proc. WWW*, 2011.

[33] J. Teevan, S. T. Dumais, and D. J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In *Proc. SIGIR*, 2008.

[34] J. Teevan, D. J. Liebling, and G. Ravichandran Geetha. Understanding and predicting personal navigation. In *Proc. WSDM*, 2011.

[35] Y. Ustinovskiy and P. Serdyukov. Personalization of web-search using short-term browsing context. In *Proc. CIKM*, 2013.

[36] Y. Ustinovskiy, G. Gusev, and P. Serdyukov. An optimization framework for weighting implicit relevance labels for personalized web search. In *Proc. WWW*, 2015.

[37] L. Wang, P. N. Bennett, and K. Collins-Thompson. Robust ranking models via risk-sensitive optimization. In *Proc. SIGIR*, 2012.

[38] Q. Wu, C. Burges, K. Svore, and J. Gao. Adapting boosting for information retrieval measures. *Journal of Information Retrieval*, 2009.

[39] B. Xiang, D. Jiang, J. Pei, X. Sun, E. Chen, and H. Li. Context-aware ranking in web search. In *Proc. SIGIR*, 2010.