
Appendix: A Comparison between Predictive and Projective Ordering

Lemma 2.1 shows that when only a single list of ads is precomputed and used for serving, and the goal is to maximize the probability of retrieving the highest-scoring ad as quickly as possible, then it is optimal to order the ads in the list by their probability of being the top ad. Thus, when a single list is allowed, the predictive approach to ordering the ads is the correct procedure. In this section, we show that when more than one list is allowed, the projective approach may be superior to the predictive approach.

We proceed by modifying the example provided in Section 2.2. Suppose there are $n \geq 2$ page features. Let $P^i = (0, \dots, 0, 1, 0, \dots, 0)$ be the page, represented as a vector, that has a one in position i and a zero in all other positions. Let $\bar{P} = (1, \dots, 1)$ be the page that has a one in all positions. Suppose the distribution of pages assigns probability of $\beta > 1/n$ to \bar{P} and equal probability to the other pages P^i for $i = 1, \dots, n$ (only these $n + 1$ pages are given non-zero probability). Further, suppose the scoring rule is linear, of the form $f(P, a) = \sum_{i=1}^n P_i g_i(a)$. Suppose there are $n + 1$ ads a_1, \dots, a_n and \bar{a} . For $i = 1, \dots, n$ let $g_i(a_j) = 1$ if $i = j$, $g_i(a_j) = -(n - 1)^{-1}$ if $i \neq j$, and $g_i(\bar{a}) = \alpha$. Thus, for all i and $j \neq i$, we have that $f(P_i, a_i) = 1$, $f(P_i, a_j) = -(n - 1)^{-1}$, $f(\bar{P}, a_i) = 0$, $f(P_i, \bar{a}) = \alpha$, and $f(\bar{P}, \bar{a}) = n\alpha$. We assume that $0 < \alpha < 1$, so in particular, $1 = f(P_i, a_i) > f(P_i, \bar{a}) = \alpha$ and $f(\bar{P}, \bar{a}) = n\alpha > 0 = f(\bar{P}, a_i)$. Therefore, the projective method will always place ad a_i first in the list associated with feature i .

Now, consider the predictive approach that uses the standard feature covering (see Section 3.1). For list i , the probability that ad \bar{a} is the top, over the pages that have feature i , is proportional to β and is greater than the probability that any other ad is top. Thus, suppose we run the predictive and projective approaches and allow the algorithms to fully evaluate only one ad per serve. Then, the predictive method will always chose to display ad \bar{a} , and it will be right β fraction of the time. On the other hand, the projective approach will choose ad a_i for page P^i , and will be right $1 - \beta$ fraction of the time. Thus, the projective approach is correct $(1 - \beta)/\beta$ times as much as the predictive approach, and is infinitely better as $n \rightarrow \infty$. The same argument holds when we allow k evaluations provided we make k copies of ad \bar{a} . In addition, the same argument holds if the predictive approach orders by expected score (rather than the probability of being the top ad) as long as we set $\alpha \geq \beta$.