Batch-Mode Computational Advertising based on Modern Portfolio Theory

Dell Zhang and Jinsong Lu

Birkbeck, University of London
London WC1E 7HX, UK
dell.z@ieee.org, jingsong.lu@gmail.com

Abstract. The research on computational advertising so far has focused on finding the single best ad. However, in many real situations, more than one ad can be presented. Although it is possible to address this problem myopically by using a single-ad optimisation technique in serial-mode, i.e., one at a time, this approach can be ineffective and inefficient because it ignores the correlation between ads. In this paper, we make a leap forward to address the problem of finding the best ads in batch-mode, i.e., assembling the optimal set of ads to be presented altogether. The key idea is to achieve maximum revenue while controlling the level of risk by diversifying the set of ads. We show how the Modern Portfolio Theory can be applied to this problem to provide elegant solutions and deep insights.

1 Introduction

Online advertising has become a major industry. It is now an important source of income for many Web sites, particularly search engines such as Google and Yahoo!.

The research on computational advertising so far has focused on finding the *single* best ad [1]. However, in many real situations, more than one ad can be presented. For example, both Google and Yahoo! currently display up to 8 ads (sponsored links) for each query. Although it is possible to address this problem myopically by using a single-ad optimisation technique in *serial-mode*, i.e., one at a time, this approach can be ineffective and inefficient because it totally ignores the correlation among the "best" ads. While the selected ads all have high expected revenue, they can be very similar to each other, therefore displaying those ads is like "putting all eggs in one basket".

In this paper, we make a leap forward to address the problem of finding the best ads in *batch-mode*, i.e., assembling the optimal set of ads to be presented altogether. Our approach to batch-mode computational advertising is motivated by two observations: (1) the future revenues of ads are inherently uncertain; (2) the future revenues of ads are usually correlated with each other. The key idea is to achieve maximum revenue while controlling the level of risk by *diversifying* the set of ads. For example, given the query 'London weather', even if the most profitable ads are all from companies selling umbrellas, it could be a better

strategy for the search engine to show a mixture of ads from umbrella companies and sunscreen companies, because the revenue would be more stable. For another example, given the query 'fashion magazine', men and women are probably looking for different products, therefore displaying some ads for men and some ads for women would give every user something relevant no matter what the gender is, and thus provide a better user experience overall and hopefully lead to an increase in revenue.

2   Approach

Assume that there are n ads $a_1, a_2, \ldots, a_n$ available in the advertising system. Given k ad places in the target Web page (either a search result page in 'sponsored search' or a content page in 'content matching'), the problem of batch-mode computational advertising is to select the optimal set of k ads.

We think this problem can be recast in the language of investment as follows. Each ad a_i is an *asset* (e.g., stock) with future *return* $r_i \in \mathbb{R}$ $(i = 1, \ldots, n)$, which is determined by its bid price and click-through rate (CTR) following the popular pay-per-click (PPC) model. The CTR of each ad can be estimated from the historical data or approximated by the relevance of the ad to the query or the contextual page. The future return on a risky asset is inherently uncertain, so r_i should be regarded as a random variable. Suppose that the mean of r_i is $E(r_i) = \mu_i$ and the variance of r_i is $Var(r_i) = \sigma_i^2$. Moreover, let σ_{ij} be the covariance between r_i and r_j for all $1 \leq i, j \leq n$. It is well-known that $\sigma_{ii} = \sigma_i^2$, and when $i \neq j$ we have $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ where $\rho_{ij} \in [-1, 1]$ is the correlation coefficient between r_i and r_j. The covariance values can be estimated from the historical data or approximated using the pair-wise similarity of ads.

A set of k ads, S, can be considered as a *portfolio* of assets. Let a binary variable $b_i \in \{0, 1\}$ indicate whether a_i is selected: $b_i = 1$ if $a_i \in S$ or 0 otherwise. Let $w_i = b_i/k$, i.e., the fraction of the ad a_i in the portfolio. Then the overall future return of the portfolio, $r_p = \sum_{i=1}^{n}(b_i r_i)/k = \sum_{i=1}^{n} w_i r_i$, is characterised by its mean and variance: $E(r_p) = \sum_{i=1}^{n} w_i \mu_i$, $Var(r_p) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}$.

We would like to find the optimal portfolio that has the maximum return for a given risk, or equivalently the minimum risk for a desired return $E(r_p) = \mu_p$ of portfolio. Here the *risk* of portfolio is quantified by $Var(r_p)$: the less variance, the less volatility, the less risk.

$$\text{minimize} \quad \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}$$

$$\text{subject to} \quad \sum_{i=1}^{n} w_i \mu_i = \mu_p$$

$$\sum_{i=1}^{n} w_i = 1$$

$$\forall 1 \leq i \leq n : w_i k \in \{0, 1\}$$

However, it turns out that the above combinatorial optimisation problem is NP complete and thus computational intractable. Therefore we relax the constraint to allow $w_i = b_i/k$ to take any real value in $\mathbb{R}$. The value of w_i can be considered as the weight of ad a_i. We first solve the following continuous optimisation problem to get the optimal weights, and then select the top k ads with highest weights as an approximation of the optimal portfolio.

$$\text{minimize} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij}$$

$$\text{subject to} \quad \sum_{i=1}^{n} w_i \mu_i = \mu_p$$

$$\sum_{i=1}^{n} w_i = 1$$

In addition to making the computation feasible, we are now able to apply Modern Portfolio Theory (MPT) [2] to this problem to get elegant solutions and deep insights.

We can rewrite the above problem in matrix-vector form as follows:

$$\text{minimize} \quad f(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathbf{V}\mathbf{w}$$

$$\text{subject to} \quad g_1(\mathbf{w}) = \mathbf{w}^T \mathbf{e} - \mu_p = 0$$

$$g_2(\mathbf{w}) = \mathbf{w}^T \mathbf{1} - 1 = 0 \ ,$$

where $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$, $\mathbf{e} = (\mu_1, \mu_2, \ldots, \mu_n)^T$, $\mathbf{1} = (1, 1, \ldots, 1)^T$, and $\mathbf{V} \in \mathbb{R}^{n \times n}$ is the covariance matrix with $\mathbf{V}(i, j) = \sigma_{ij}$, $1 \le i, j \le n$. Using Lagrange multipliers, we can solve the above problem analytically to get the optimal vector of portfolio weights

$$\mathbf{w}_p = \frac{1}{D}(B\mathbf{V}^{-1}\mathbf{1} - A\mathbf{V}^{-1}\mathbf{e}) + \frac{1}{D}(C\mathbf{V}^{-1}\mathbf{e} - A\mathbf{V}^{-1}\mathbf{1})\mu_p \ ,$$

where $A = \mathbf{e}^T\mathbf{V}^{-1}\mathbf{1}$, $B = \mathbf{e}^T\mathbf{V}^{-1}\mathbf{e}$, $C = \mathbf{e}^T\mathbf{V}^{-1}\mathbf{1}$, and $D = BC - A^2$.

The above analytical solution is helpful in understanding the optimal portfolio, but it is computational expensive as it involves inversion of a dense matrix $\mathbf{V}$. In practice, we can use numerical computation techniques to get the numerical solution efficiently.

Every possible portfolio can be plotted in the risk-return space (with return μ_p on the y-axis and risk σ_p on the x-axis), and the collection of all such portfolios defines a region in this space. The hyperbola along the upper edge of this region is known as the *efficient frontier* (aka the Markowitz frontier), as illustrated in Fig 1. Combinations along this line represent portfolios for which there is lowest risk for a given level of return. Conversely, for a given amount of risk, the portfolio lying on the efficient frontier represents the combination offering the best possible return. The efficient frontier is the set of portfolios for which
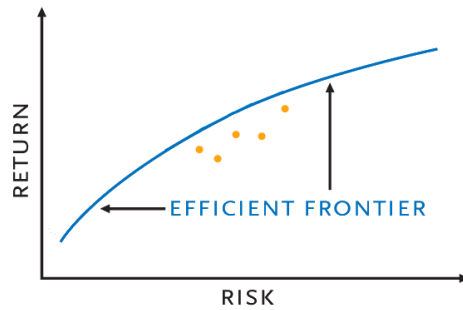
Fig. 1. The efficient frontier.

one cannot improve both risk and return. On one hand, the region above the efficient frontier is unachievable by holding risky assets alone, i.e., no portfolios can be constructed corresponding to the points in this region. On the other hand, points below the frontier are suboptimal. Therefore a rational investor will hold a portfolio only on the frontier.

3 Conclusions

This paper presents a sketch theoretical development towards batch-mode computational advertising based on Modern Portfolio Theory (MPT). It is necessary to perform large scale experiments on real-world ad datasets to empirically evaluate our proposed approach, and compare it with existing heuristic methods in information retrieval for diversifying search results (such as MMR [3]). Furthermore, due to the sparsity of ad click-through data, how to estimate the future return of ads and their correlations effectively and efficiently remains to be an open research problem.

Acknowledgements

References

1. Radlinski, F., Broder, A.Z., Ciccolo, P., Gabrilovich, E., Josifovski, V., Riedel, L.: Optimizing relevance and revenue in ad search: A query substitution approach. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Singapore (2008) 403–410
2. Markowitz, H.M.: Portfolio selection. Journal of Finance **7**(1) (1952) 77C–91
3. Carbonell, J.G., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Melbourne, Australia (1998) 335–336