Dynamical Information Retrieval Modelling

A Portfolio-Armed Bandit Machine Approach

Marc Sloan and Jun Wang
Department of Computer Science
University College London
{M.Sloan, J.Wang}@cs.ucl.ac.uk

ABSTRACT

The dynamic nature of document relevance is largely ignored by traditional Information Retrieval (IR) models, which assume that scores (relevance) for documents given an information need are static. In this paper, we formulate a general Dynamical Information Retrieval problem, where we consider retrieval as a stochastic, controllable process. The ranking action continuously controls the retrieval system's dynamics and an optimal ranking policy is found that maximises the overall users' satisfaction during each period. Through deriving the posterior probability of the documents evolving relevancy from user clicks, we can provide a plug-in framework for incorporating a number of click models, which can be combined with Multi-Armed Bandit theory and Portfolio Theory of IR to create a dynamic ranking rule that takes rank bias and click dependency into account. We verify the versatility of our algorithms in a number of experiments and demonstrate improved performance over strong baselines and as a result significant performance gains have been achieved.

Categories and Subject Descriptors: H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval

General Terms: Algorithms, Experimentation, Measurement, Performance

Keywords: Multi-period information retrieval, portfolio theory, statistical modeling, multi-armed bandit machine

1. INTRODUCTION

Mainstream IR models and theories have been largely devoted to maximising static performance, however, the relevance of a document is in fact stochastic and liable to change over time. The *Dynamical Information Retrieval Problem* is expressed as: *"We have an IR system receiving queries over a period of time $t = 1 \dots T$, returning a set of ranked documents, and observing user clickthroughs. How do we determine an optimal ranking policy that maximises some utility and responds to changing document relevancies over t?"*

Our study in this paper is a theoretical one; we offer a general solution framework by using an *optimal control* formulation [1]. A dynamically controlled system requires a control signal, our rank action at time t, whose state denotes the system's belief about the documents' relevancy, and a sensor (clickthroughs) so the system can monitor it's changing environment. Interestingly, we also find that the objective function can be conveniently decomposed into two parts, the mean and the variance, where the mean handles rank bias and the variance tackles click dependency. Traditionally, IR systems work under the Probability Ranking Principle (PRP), whereas we can incorporate the portfolio theory of IR [5] to extend the ranking principle to address the dependency issue and promote diversity.

We address three critical issues: 1) The system's state is the posterior probability of a document being relevant, and we propose an iterative update mechanism 2) The system's dynamic function is derived, which shows how the posterior evolves according to past rank decisions and user feedback 3) We demonstrate use of the dynamic function using simple ranking rules, integrating portfolio ranking with multi-armed bandit machine research [2]. Simulated experiments confirm the theoretical insights with improved performance in addressing rank bias and dependency of user feedbacks.

2. MODEL AND ALGORITHMS

In our optimal control formulation, our objective is to maximise the expected number of clicks R_t over time $\sum_{t=1}^{T} E[U(R_t)]$, using the exponential utility

$$U(R_t) = 1 - \exp(-\lambda \sum_{t=1}^{T} R_t)$$

Our optimal rank action $\mathbf{a}^*$ is then that which maximises this expected value, which can be rewritten as

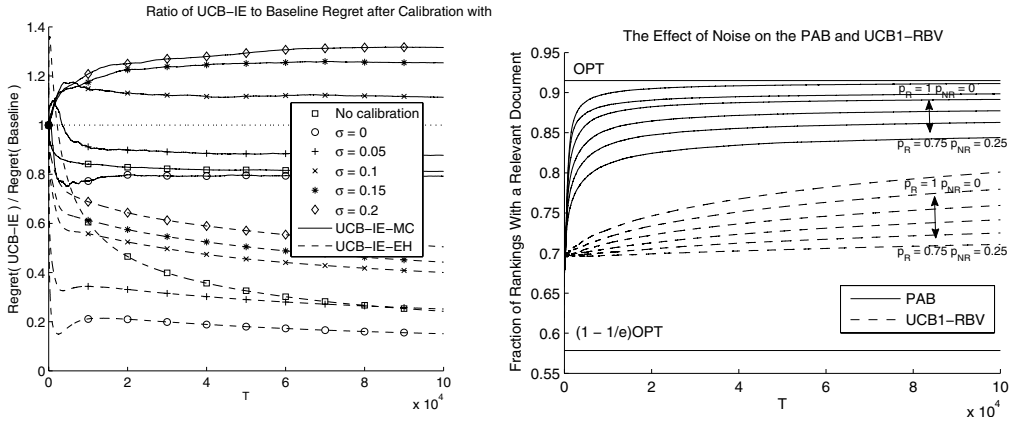
$$\mathbf{a}^* = \underset{(a(1)\dots a(T))}{\operatorname{argmax}} \sum_{t=1}^{T} (E[R_t] - \frac{\lambda}{2}\mathrm{Var}[R_t])$$

Firstly, our Iterative Expectation (UCB-IE) algorithm ignores the variance (by setting $\lambda = 0$), and models R_t using a *click model*, given by

$$E[R_t] = \sum_{i=1}^{M} p(C = 1 | D = d_i, i)$$

where d_i is the ith ranked document, M the number shown to the user and C the binary event of a click. To incorporate rank bias, we assume a mixture click model [3] which introduces a binary latent variable S_i, indicating whether a user clicks due to the bias of rank i or due to relevance. This general click model allows us to introduce the parameters $r_d = p(C = 1|d_i)$, $b_i = p(C = 1|S_i = 0)$ and $\pi_i = p(S_i = 0)$,

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.



(a) Ratio of UCB-IE and baseline regrets (b) Effect of noise on PAB and baseline

Figure 1: Performance of UCB-IE and Portfolio Bandit algorithms

respectively, the probabilities of document relevance, clicking due to bias and clicking due to relevance. Setting these parameters allows us to plug-in specific click models, and in our experiment we implement the mixed clicks and examination hypothesis models [3].

By making use of an Expectation Maximization algorithm and assuming that documents are independent, we can simplify the model and derive an iterative update formula. This formula allows us to update our estimated probability of relevance $\hat{r}_d$ by using the model parameters to determine 'psuedo-counts' for the events of both observing and not observing a click, and thus is our dynamical solution to this problem.

This formula can be plugged into a UCB style algorithm that can dynamically learn and exploit the probabilities of relevance for a series of documents. At each time step, documents are ranked and the top M displayed according to an index $\Lambda_d = \hat{r}_d + \sqrt{\frac{2 \ln t}{B_d(t)}}$, where B_d is considered the *effective* number of times the document has been displayed. After observing if the document was clicked, B_d is updated using its previous value and the model parameters, and then used to update a new estimate for $\hat{r}_d$.

For the Portfolio-Armed Bandit (PAB) algorithm, we allow $\lambda > 0$ and take into account document dependence and diversity. The ranking index this time is $\Lambda_d = \frac{X_d(t)}{Y_d(t)} + \sqrt{\frac{2 \ln t}{Y_d(t)}}$, where $X_d(t)$ and $Y_d(t)$ are the number of times d has been clicked and displayed respectively up until t. We learn the *correlations* between documents by observing their co-clicks, the number of times two documents are clicked together in the same ranking, and use this to re-rank them using portfolio theory [5].

3. EXPERIMENTS AND CONCLUSION

Owing to the difficulty in evaluating dynamic algorithms using offline search logs, appropriate simulations were used for experimentation. We compared two UCB-IE click model variants, mixed clicks (MC) and examination hypothesis (EH), against a click model agnostic UCB baseline, and measured the expected click regret, calculated as the accumulated difference between an optimal ranking and the one chosen by the algorithm.

After observing a statistically significant improvement over the baseline for both algorithms, we introduced a calibration phase, simulating how a live system could first learn relevancies using pre-computed relevance scores (such as BM25), and then adjust to a 'real-world' setting which may have subtly different scores. Fig. 1a shows how well each algorithm adjusted after the calibration phase for different levels of 'real-world' variation σ, where curves underneath the dotted horizontal line indicate improved performance over the baseline, with UCB-IE-EH performing well but UCB-IE-MC sometimes worse than the baseline.

The PAB algorithm was tested on a simulation containing 50 documents belonging to randomly assigned topics and scored according to the diversity of a ranking. It was compared to the UCB1 Ranked Bandits Variant (UCB1-RBV) [4] and showed significant improvement for various values of λ (our risk preference). Settling on a value of $\lambda = 1$, we then added increasing levels of noise to both algorithms and Fig. 1b shows that not only did the PAB consistently perform well, but was also resistant to noise.

We have proposed a theoretically robust, non-static ranking solution to our dynamical information retrieval problem, that can take into account rank bias and diversity. We have implemented the solution in a number of example algorithms, which we then demonstrate to perform well in a simulated setting. We intend to conduct experiments using search data, whilst also exploring further the theoretical justification of the algorithms, proving their effectiveness, and also try to alleviate some of the strong assumptions.

4. REFERENCES

[1] Athans, M., and Falb, P. *Optimal Control: An Introduction to the Theory and Its Applications*. Dover Publications, 2006.

[2] Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn. 47* (2002).

[3] Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. An experimental comparison of click position-bias models. WSDM (2008), ACM.

[4] Radlinski, F., Kleinberg, R., and Joachims, T. Learning diverse rankings with multi-armed bandits. ICML (2008), ACM.

[5] Wang, J., and Zhu, J. Portfolio theory of information retrieval. SIGIR (2009), ACM.