

**Efficient Parameter Estimation for Information Retrieval using Black-box Optimization**

**Abstract:**

The retrieval function is one of the most important components of an Information Retrieval (IR) system, because it determines to what extent some information is relevant to a user query. Most retrieval functions have “free parameters” whose value must be set before retrieval, significantly affecting the effectiveness of an IR system. Choosing the optimum values for such parameters is therefore of paramount importance. However, the optimum can only be found after a computationally expensive process, especially when the generalization error is estimated via cross-validation. In this paper, we propose to determine free parameter values by solving an optimization problem aimed at maximizing a measure of retrieval effectiveness. We employ the black-box optimization paradigm, since the analytical expression of the measure of effectiveness with respect to the free parameters is unknown. We consider different methods for solving the black-box optimization problem: a simple grid-search over the whole domain, and more sophisticated techniques such as line search and surrogate model based algorithms. Experimental results on several test collections not only provide useful insight about effectiveness, but also about efficiency: they indicate that with appropriate optimization techniques, the computational cost of parameter optimization can be greatly reduced without compromising retrieval effectiveness, even when taking generalization into account.

**Existing System:**

Parameters are typically optimized only at the time of construction and indexing of the document collection, but optimization should be performed whenever document collections grow, or queries evolve from one style to the other: for example, three query styles – informational, navigational and transactional – were found in the context of theWeb.

Therefore, the overall effectiveness of a retrieval system is harmed by the inefficiency of performing repeated parameter optimization for every situation, i.e., collection or query set. The problem is that the degree of inefficiency grows exponentially with the number of parameters, since dividing n parameters into k subranges would require testing O(kn) different configurations. This can be very inefficient for as little as three parameters. In fact, studies in machine learning reveal that even random search (testing several random values sampled from a uniform distribution) can work better than grid search

**Proposed System:**

The main contribution of this paper consists in casting the parameter optimization problem as a mathematical program solved by black-box optimization methods that do not require an analytical description of the objective function, and in empirically showing that (i) black-box optimization is significantly more efficient than grid search on a benchmark set, and (ii) blackbox optimization reliably and efficiently (i.e., in a short time) finds effective parameter values even when generalization is a requirement and/or the retrieval function has many parameters.