**Pseudo-Relevance Feedback Using Text Classification**

1 INTRODUCTION  
2 PRIORWORK

The idea of treating document ranking as a binary classification problem, to distinguish relevant from non-relevant documents, has a long history, dating back to the binary independence retrieval (BIR) model of Robertson and Spark Jones [19].

Today, the effectiveness of pseudo-relevance feedback is well established in the literature. Note that while there is literature on how to select good expansion terms using supervised machine

learning techniques [4], my approach is completely unsupervised.

3 EXPERIMENTAL SETUP

My proposed technique has three parameters: r , the number of pseudo-positive labels, n, the number of pseudo-negative labels, and α, the interpolation weight.

4 RESULTS

Across all four collections, logistic regression in most cases is slightly better than SVMs in terms of effectiveness, but the gains from logistic regression are not significant for all collections. Results further suggest that an ensemble based on simple score averaging yields no benefit over individual models.

5 DISCUSSION