

Genetic Algorithm-Based Query Expansion for Improved Information Retrieval

Pragati Bhatnagar and Narendra Pareek

Abstract This paper is focused toward query expansion, which is an important technique for improving retrieval efficiency of an information retrieval system. In particular, the paper proposes an evolutionary approach for improving efficiency of pseudo-relevance feedback-based query expansion (PRFBQE). In this method, the candidate terms for query expansion are selected from an initially retrieved list of documents, ranked on the basis of co-occurrence measure of the terms with the query terms. Top n selected terms create a term pool. From this term pool, genetic algorithm (GA) is used to select a thematically rich combination of terms, which provides the terms for expanding the query. We call this method as genetic algorithm-based query expansion (GABQE). The experiments were performed on standard CISI dataset. The results are quite motivating, and one can clearly observe the difference in the result when GA is not used and when GA is used. The paper uses GA for improving pseudo-relevance feedback (PRF)-based query expansion, but at the same time, it can also be generalized and tested for other types of query expansions, where terms may be selected in a different way, but a good combination of expansion terms can be obtained using GA.

Keywords Information retrieval • Query expansion • Genetic algorithm

1 Introduction

Query expansion has been widely investigated as a method for improving the performance of information retrieval system. Though a lot of work has been done in this area, obtaining a proper expansion of query is still an unsolved problem. Different researchers are coming up with different techniques of query expansion.

P. Bhatnagar (✉) · N. Pareek
Department of Computer Science, M.L. Sukhadia University, Udaipur, Rajasthan, India
e-mail: pragatibhat@gmail.com

© Springer India 2015
L.C. Jain et al. (eds.), *Intelligent Computing, Communication and Devices*,
Advances in Intelligent Systems and Computing 308,
DOI 10.1007/978-81-322-2012-1_6

47