Summary

Several algorithms that allow user interaction with an automatic document retrieval system by requesting relevance judgments of selected sets of documents are investigated. All relevance feedback algorithms tested improve the average retrieval obtained. The improvement caused by relevance feedback is greater in a larger, more realistic document collection (Cranfield 200) than in a smaller and less realistic collection (ADI).

No single feedback strategy is found to give superior retrieval for all queries. Algorithms using only relevant documents for feedback can be made effective for queries that retrieve no relevant documents on the initial search in two ways, by supplying additional documents for relevance judgments or by using non-relevant documents for feedback.

In general, the performance of negative feedback algorithms (those using non-relevant documents for feedback) is variable. Negative feedback gives worse performance on some queries and better performance on others than the more consistent positive feedback strategies. On the average, negative feedback moves the query closer to the optimum query defined by Rocchio than does positive feedback, but does not differ significantly from positive feedback from the viewpoint of the user during feedback.

A study of selected subgroups of queries provides three

interesting results. First, movement of the query constructed by the Rocchio strategy further away from the original query rather than back toward it on the second iteration is related to poor performance and poor initial search results. This direction of query movement could be a result of inadequate feedback, or it could be an attempt to compensate for a poor original query. Second, the queries for which the Rocchio strategy gives less improvement than positive feedback have the worst initial search performance, queries for which positive feedback is inferior have much better initial search performance, and queries for which negative and positive feedback are equal have the best initial search retrieval. Third, queries having few concepts and few relevant documents and queries having many concepts and few relevant documents tend to give less improvement with negative feedback than with positive feedback, while queries for which the number of concepts and number of relevant documents are directly related give more improvement with negative feedback. If a predictor of the number of relevant documents available for each submitted query can be found, this relationship could be used to select the feedback algorithm appropriate to each query.

The observed contrasting behavior of negative and positive feedback algorithms can be explained by a hypothesis presented in Section VII-C.

Hypothesis:

A hypothesis is presented that explains some of the observed

performance differences between the negative feedback strategies and the positive feedback strategies investigated, and is consistent with all experimental results reported.

Hypothesis:

For most queries, for every vector v contained in the set R of relevant document vectors there exists at least one vector s contained in the set S of non-relevant document vectors such that for some other vector r contained in R, cos(r,s) is greater than cos(v,r). Further, for a significant number of queries the prevalence of such relationships effectively prevents the retrieval of some relevant documents with reasonable precision by any relevance feedback strategy.that constructs only one query on each iteration.

This hypothesis states in effect that the documents relevant to a single query are usually found in two or more distinct clusters in the concept vector space, and that these clusters of relevant documents are separated from each other by non-relevant documents. Further, it states that for a significant number of queries this phenomenon will seriously interfere with the retrieval of some relevant documents regardless of the relevance feedback strategy employed. For any collection in which this hypothesis is true, all relevance feedback algorithms tested in this study are inappropriate for a significant percentage of retrieval requests. Algorithms constructing more than one query on each feedback iteration are necessary in such an environment.

The anomalous results of the reported comparisons of positive

and negative feedback support the conclusion that the stated hypothesis is true in the Cranfield 200 collection. Because this collection is a carefully chosen subset of a larger collection representative of a well-defined, technical, limited subject area, this conclusion suggests that multiple query algorithms or other means of simplifying the distribution of relevant document vectors in the vector set being searched will be needed in practical automatic retrieval systems.

Section VII contains many recommendations for future research in relevance feedback. Positive feedback with the combination feedback strategy that presents more documents to users that judge no documents relevant on a given iteration is recommended in Section VII-A for environments similar to the Cranfield 200 collection. Non-relevant document feedback is recommended for users who find no relevant documents after a maximum number of additional documents have been presented. Section VII-B discusses the evaluation problems encountered in this study. New global measures similar to normalized recall and normalized precision are suggested. The Quasi-Cleverdon interpolation method is recommended in preference to Neo-Cleverdon interpolation for recall-precision curves. Three new evaluation viewpoints for relevance feedback are suggested, one of which is appropriate to other areas. A fourth evaluation method is discussed and recommended for general use.

Section VII-D discusses the implications of the conclusion reached from the hypothesis of Section VII-C for partial search strategies, multiple query strategies, request clustering, and document space modification. Two new measures for evaluating the usefulness of a given partition of the document collection regardless of the partial search algorithm employed are presented. The cluster search algorithm used in earlier studies of Rocchio's clustering algorithm in the SMART system is inap

is inappropriate where groups of documents relevant to the same query are separated by non-relevant documents, so a new cluster search algorithm is presented. This new algorithm is combined with relevance feedback to form a cluster feedbackalgorithm that creates a different query to search each document cluster. Several design considerations for multiple query algorithms, strategies that construct more than one query to search the same set of documents, are proposed, culminating in a detailed algorithm simple enough to be meaningfully tested in a small document collection. Finally, request clustering and document space modification are discussed as possible ways of making multiple query algorithms unnecessary by additional processing that does not take place during the search. An algorithm for permanent adaptive alteration of the document vectors using queries and relevance judgments is constructed by analogy to a well-tested algorithm that performs a similar function in adaptive pattern recognition systems.