Filters and Answers: The University of Iowa TREC-9 Results

Elena Catona[1], David Eichmann[2,3] and Padmini Srinivasan[1,2]

[1]Department of Management Sciences
[2]School of Library and Information Science
[3]Computer Science Department
The University of Iowa
{elena-catona, david-eichmann, padmini-srinivasan}@uiowa.edu

The University of Iowa participated in the adaptive filtering and question answering tracks of TREC9. The filtering system used was an extension of the one used in TREC-7 [1] and TREC-8 [2]. Question answering was done using a rule-based system that employed a combination of public domain technologies and the SMART retrieval system.

1 – Adaptive Filtering:

Our approach to filtering involves a two-level dynamic clustering technique. Each filtering topic is used to create a primary cluster that forms a general profile for the topic. Documents that are attracted into a primary cluster participate in a topic-specific second level clustering process yielding what we refer to as secondary clusters. These secondary clusters, depending upon their status, are responsible for declaring, i.e., retrieving, documents for the topic.

As documents are temporally processed they are attracted to a primary cluster if their similarity with the cluster vector is above a primary threshold. These documents enter the secondary clustering stage where again, based on similarity to cluster vectors and a secondary threshold, they either join an existing secondary cluster or start a new one. If at some point the similarity between a secondary cluster and the primary cluster exceeds a third declaration threshold then the document most recently added to the secondary cluster is retrieved for the user.

When deriving representations we use TF*IDF weights after stemming the terms using Porter's stemmer. We also limit document vectors and cluster vectors to the best 100 and 200 stems respectively.

In TREC-8 adaptation was explored at several different levels [2]. First a secondary cluster's future behavior would depend upon past performance. If a secondary cluster declares a document that turns out to be relevant then it is colored green. This means that it declares all documents that join it in the future. If instead the declared document is non relevant then the cluster is colored red and all future documents are not declared. A non relevant document that joins a green cluster spawns an independent red cluster allowing the original cluster to remain green. Another adaptive dimension was to have the primary cluster vector adapt as relevant judgements were obtained. A version of Rochio's feedback method is built into the system for this purpose. A differential adaptation scheme is also built in for this purpose. The key distinction is that in the differential scheme positive and negative term vectors are comprised only of terms not found in the other vector or in the original query vector.

Recent experiments conducted with TREC-8 data explored additional dimensions of adaptation. For example, we experimented with adapting the primary threshold as the performance measure varied. For this, the performance measure such as the utility score was computed at regular intervals when a "snapshot" of the system is taken. We also explored adaptation of secondary and declaration thresholds. In all these the most profitable approach appears to be adaptation of the break threshold using a step function that responds to changes in performance across snapshots. Our OHSU runs use the system as described above with adaptation of the break threshold.

Other key extensions to the system for TREC-9 include the ability to specify the type of index vectors to utilize. A phrase recognizer loads dictionaries of phrases derived from sources such as the WordNet thesaurus and matched phrases are included into the document vectors. More recently a rule-based entity recognizer has been developed that allows the indexing of documents by person names, organizations, locations and events. Our MESH run includes this technology as well special support for medical terminology. The MeSH hierarchy, an associated lexicon of synonyms and a supplementary list of concepts such as drug names were used. The MESH run involved index vectors that were populated using only the entities extracted from the source text.

OHSU Runs:

For TREC-9 we submitted two OHSU runs. These runs employed word based indexing, Rochio feedback for the profile adaptation and adaptation of the declaration threshold. Both OHSU runs used the controlled vocabulary field (MeSH terms). The two runs differ only in their starting threshold values. The primary, secondary and declaration thresholds were 0.3, 0.32, and 0.3 respectively for OHSU1 and 0.25, 0.27, and 0.25 respectively for OHSU2. The declaration threshold was adapted in each case using a step-wise strategy. Figure 1 shows the performance in terms of utility for our OHSU1 run. The dashed bars represent median performance across systems for each topic. There are 24 topics for which OHSU1 was better than the median and another 24 for which it was below the median.

We conducted several experiments after the official submission deadline to better understand the different aspects of our filtering system and its weak performance on the OHSU task. The first question asked was whether the primary filter was effective. In other words how good was it at filtering out non relevant documents while allowing through the relevant documents? Figure 2 shows the percentages filtered through over time, with snapshots taken every 1000 documents. The figure shows that if we divide the snapshots into three groups then the primary filter allows about 50%, 60% and then 59% of the relevant documents that arrive over the first, second and third sequence of snapshots respectively. At the same time the percentage of non relevant documents allowed through stays less than 1% of the number seen. We then examined the effectiveness of the secondary filter. Note that this analysis of the secondary filter was limited to those documents allowed through by the primary filter. Figure 3 shows that the secondary filter was successful in reducing the percentage of non relevant documents allowed through (dashed bars). However, at the same time it also restricts the passage of relevant documents although not as severely. Next we took a different track in our analysis and examined the effectiveness in adapting the declaration threshold. Figure 4 displays these results. We can observe that if we eliminate break threshold adaptation performance degrades significantly over time (dashed bars). In contrast, the adaptive mode is able to

stay somewhat steady - although on the negative side of the performance axis. At this point we suspected that our break threshold may not be restrictive enough. Figure 5 shows the effect of testing this by contrasting a run where the break threshold was increased from the original 0.25 (dashed bars) to 0.3 (solid bars).