Automating the Creation of Information Filters

Curt Stevens
stevens@cs.colorado.edu

One of the big probelms in using filters to deal with large information spaces is the effort involved in creating, maintaining, and evolving filters over time. The effort involved can cause enough overload to make filter management as much trouble as the overload problems filters were intended to solve. In order to address this situation the INFOSCOPE [1] system employs rule–based *agents* that recognize a user's usage patterns and make suggestions based on them. These suggestions help users create and maintain their own sets of filters.

Agents keep a constantly evolving user model of individual interests. As users read Usenet news messages, data about their interactions is stored in a knowledge base. Agents use rules to measure and record interesting terms found in read and deleted mail, as well as rules for tracking the timeliness of those measurments. Terms are collected from header fields and meaningless terms (like if, and, was...) are eliminated. Frequency and recency might be determined by ensuring terms are read in 30% or more of the last 6 sessions, or 50% of the last 10 sessions. These tests are fully adjustable by individual users and therefore the rules might be different for each user. A pattern that is frequently supported by additional readings of matching items will continue to trigger a suggestion, while patterns that are less frequently supported, or have been allowed to sit without action will be removed from the suggestion queue. Using this approach agents make suggestions that are completed filters based on observed reading patterns. They are displayed in a graphical dialog box with editable text fileds for altering the terms to be included in or excluded from the filter, as well as what newsgroups it should search. The user is in the position of filter critic instead of filter constructor. Users can define filters by editing suggestions instead of having to remember each term for manual filter creation. Terms presented in the suggestion can be removed and additional terms added. The suggestions serve as cues to remind users of what they have been reading. This can be especially helpful in situations where users do not recognize the patterns of interest they exhibit.

An excerpt from a naturalistic user study can help demonstrate how this works. Subject A is the president of a small computer consulting firm and has used other news readers for many years. The subject is very familiar with Usenet news. Almost immediately after installing INFOSCOPE on an office Macintosh this user begins defining filters for interesting terms from the rec.bicycles, comp.sys.sun.admin. and comp.protocols.ppp newsgroups. Within a week, however, usage data shows that an additional newsgroup (misc.forsale.computers) is being read several times. A frequency pattern for that newsgroup is added to the knowledge base. Over the next week agents gather terms from messages read in this newsgroup. These terms include the words "sale," "Imagewriter," "Macintosh," and "Mac." Each time the user supports the frequency pattern by reading more from the newsgroup, the weight of that pattern is increased. Within a few more sessions the frequency pattern is triggered and checked for recency. Since the patterns are very recently supported, it is then made into a suggestion. Users can ask for

explanations of why a suggestion was triggered, and modify the triggering conditions if necessary. This suggestion was modified by the user who removed the terms “sale,” “Macintosh,” and “Mac” from the filter and added the terms “cx,” and “ci” (two specific Macintosh models). The suggestion was accepted after modification, creating the newsgroup
misc.forsale.computers.interesting.

INFOSCOPE agents are not perfect, often including extraneous terms that must be edited out of suggested filters. However, while INFOSCOPE agents cannot predict all user intersts, they have proven to be an effective method for bringing unknown interest patterns to the attention of news readers.

1. Fischer, G. and Stevens, C. Information Access in Complex, Poorly Structured Information Spaces. In *Human Factors in Computing Systems, CHI’91 Conference Proceedings* (New Orleans, LA, April, 1991), 63-70.