MailCat: An Intelligent Assistant for Organizing E-Mail

Richard B. Segal and Jeffrey O. Kephart
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
rsegal@watson.ibm.com, kephart@watson.ibm.com

Abstract

MailCat is an intelligent assistant that helps users organize their e-mail into folders. It uses a text classifier to learn each user's mail-filing habits. MailCat uses what it learns to predict the three folders in which the user is most likely to place each incoming message. It then provides shortcut buttons to file each message into one of these three folders. When one of MailCat's predictions is correct, the effort required to file a message is reduced to a single button click.

Introduction

Most mail readers allow users to organize their messages into folders to ease later retrieval. One might suppose that the effort required to file a message using these programs would be negligible. In practice, however, many users find the cognitive burden of deciding where to file a message plus the time spent interacting with the user interface to be a substantial barrier. This barrier is significant enough that many users quickly fall behind and let unfiled messages pile up in their mailboxes.

MailCat is an intelligent personal assistant that reduces the cognitive burden and the time required for organizing e-mail into folders (Segal and Kephart 1999). Using a text classifier that adapts dynamically to a user's mail-filing habits, it predicts the three folders that are most likely to be appropriate for a given message and provides shortcut buttons that facilitate filing that message into one of its predicted folders. When one of the folders predicted by MailCat is correct, the user's task is greatly simplified. Rather than having to derive a best choice from a large set of folders, the user can merely confirm one of MailCat's suggested choices using a single mouse click. Given that MailCat's prediction accuracy is roughly 80% to 90% even for users with as many as 60 folders, MailCat offers a qualitative enhancement that is likely to encourage users to file their mail.

MailCat offers these advantages without demanding anything in return. Users have nothing new to learn or do. When a user first installs MailCat, it analyzes their existing folders and constructs a text classifier that is tuned to that user's mail-filing behavior. Thereafter, the user may use the shortcut buttons, or may opt to use the standard message-filing interface if none of the suggested folders are appropriate. Regardless of whether its suggestions are accepted, MailCat updates its classifier promptly as each new message is filed.

MailCat

Figure 1 shows how MailCat simplifies the task of organizing messages. MailCat places three buttons above each message that allow the user to quickly file each message into one of the three folders it suggests. When one of the three buttons is clicked, the message is immediately moved to the indicated folder.

MailCat uses a text classifier to predict the likely destination folders for each message. MailCat builds its text classifier by learning from user actions. Maes (Maes 1994) suggests that it can take some time for a user to file enough messages for an e-mail assistant to learn a good classifier. Maes proposes collaborative learning as a solution to this problem in which each e-mail agent learns from other e-mail agents whose users have similar mail-filing habits. While this works well if one can find a user with similar mail-filing habits, finding such a user seems unlikely in practice given the diversity of mail-filing schemes.

An alternative solution is to learn from messages previously filed by the user. Most e-mail users already have a large database of previously-filed messages which can be used to bootstrap the text classifier — the messages currently stored in their folders. This database provides ample training data to get the classifier quickly up to speed. When MailCat is first installed, it reads the user's database of previously-filed messages and uses this information to train a TF-IDF text classifier. After this initial training, MailCat can immediately begin making useful predictions.

The initial training of the classifier is only half the battle. Users are constantly creating, deleting and reorganizing their folders. Even if the folders remain the same, the type of messages placed in a folder changes over time. MailCat adapts to changing conditions by using a classifier that supports incremental learning.

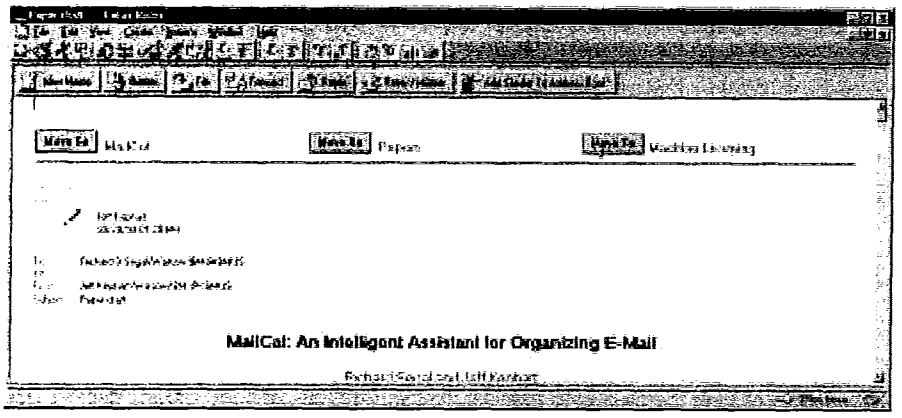

Figure 1: MailCat creates shortcut buttons for the three folders in which it predicts the user is most likely to place each message. When its predictions are correct, the user can file each message using a single button click.

Once the classifier has been trained, the classifier's model can be updated by presenting to it the messages that have been added or deleted from each folder. After updating, the classifier's predictions are indistinguishable from what they would have been had the classifier been trained from scratch on the entire mail database.

The low cost of incremental learning allows MailCat to update the classifier as the user interacts with their mail client. If the user creates a new folder and adds a few messages, MailCat instantly learns about the folder and can start predicting which messages are likely to be placed in the new folder.

While MailCat's use of three buttons rather than one may seem trivial, it is in fact quite significant because it substantially improves MailCat's usefulness. Our experimental results show that the use of three buttons cuts MailCat's failure rate in half without introducing any adverse side-effects. MailCat's use of three buttons is a direct result of our decision to provide an assistant that *facilitates* rather than *automates* message filing — a choice that itself arose from our emphasis on avoiding negative impacts on users. Since a message can be automatically filed in only one folder, automatic categorization systems have to rely solely on the accuracy of their first prediction.

Experiments

We evaluated MailCat by simulating its performance on the mailboxes of six real users. Table 1 shows the accuracy of MailCat for each user when providing from one to three shortcut buttons. The table also lists the number of folders in each user's mailbox because, as the number of folders increases, so does the difficulty of the classification problem. The accuracy of MailCat with N buttons is the frequency that one of the first N buttons it provides will move the message into the correct folder.

The results show that MailCat is fairly accurate with just one button, achieving between 59.8% and 81.1% accuracy. MailCat improves its performance by simply providing more than one shortcut button. By using three buttons, MailCat improves its accuracy to 80.2% to 98.1% — a factor of two reduction in its error rate.

User	# Folders	1	2	3
R. Segal	66	77.8	88.5	91.8
J. Kephart	56	59.8	72.8	80.2
User #3	43	64.9	78.1	84.8
User #4	34	65.9	75.7	81.1
User #5	15	70.1	88.1	93.6
User #6	14	81.1	94.4	98.1

Table 1: MailCat simulation results for six users.

Conclusions

MailCat simplifies the task of filing messages by analyzing the user's mail-filing habits to predict the three most-likely folders for each message and then providing shortcut buttons to quickly file each message into one of its predicted folders. Since its predictions are accurate 80% to 90% of the time, MailCat substantially reduces the effort required to file messages.

While MailCat was developed for electronic mail, it can easily be used to organize other types of electronic documents. The concepts behind MailCat can be applied to organizing bookmarks, files, audio recordings, and other text-based documents that are placed into a hierarchy of folders.

References

P. Maes. Agents that reduce work and information overload. *CACM*, 37(7):31–40, July 1994.

R. Segal and J. Kephart. MailCat: An intelligent assistant for organizing e-mail. In *Proceedings of Agents'99*, May 1999.