An Intelligent Interface for Sorting Electronic Mail

Elisabeth Crawford
Basser Department of Computer Science
The University of Sydney
NSW 2006 Australia
ecrawfor@cs.usyd.edu.au

Judy Kay
Basser Department of Computer Science
The University of Sydney
NSW 2006 Australia
judy@cs.usyd.edu.au

Eric McCreath
Department of Computer Science
The Australian National University
ACT 0200 Australia
ericm@cs.anu.edu.au

Abstract
Classification of email is an important everyday task for a large and growing number of users. This paper describes the **i-ems** (Intelligent-Electronic Mail Sorter) mail interface, which offers a view of the inbox based on predicted classifications of messages. The interface is designed to ensure user control over the prediction processes by supporting scrutiny of the system's certainty and details of the mechanisms used.

Keywords
Machine Learning, User Modelling, Text Classification, Inductive Logic Programming, email.

INTRODUCTION

This paper is concerned with the problem of managing email messages that the user wants to keep and the filtering of junk email. One strategy for managing email is to archive each piece into an appropriate folder. Choice of folder may depend on many factors including aspects such as the sender and nature of the email. This task is non-trivial. For example, one study of 20 workers by Whittaker and Sidner [5] found the average inbox had 2482 items. The average number of filed items was only 858. Within the study, some users left almost all their email in the inbox.

Many email programs[1] support filtering rules which can automate various mail management tasks: automated filing, deletion, replies. These rules can be expressed in terms of strings appearing in different parts of an email message. To handle an email item, the rules are evaluated in order and the first rule that applies to the item triggers the email client to move the message into the associated folder. The difficulty with rules is that the process of composing a rule is cognitively demanding and there is a real, potentially unacceptable risk of misfiling mail. Generally, users seem to avoid customising software [3][2].

In a recent study of user's management of email [2], the authors observe 'Most of our users (17 interviewees, or 60 percent) say they don't use filters. Several simply haven't figured out how to use them, suggesting that either filters need to be simpler to use or that they are not that useful.'

The aim of our **i-ems** (Intelligent-Electronic Mail Sorter) project is to investigate the automatic induction of email filtering rules. At the same time, we want to provide an intuitive interface which enables users to scrutinise both the rules constructed for them and the *reasoning underlining the system's construction of these rules*.

I-EMS INTERFACE

Direct application of a classifier might automatically sort messages into their archive folders. This is problematic for two main reasons. First, users often wish to see all the incoming mail: if messages are automatically placed in folders, important messages may be missed. Second, work is created for the user when the classifier incorrectly archives a message.

An interesting approach is taken in *MailCat* [4]. They created an interface with three buttons labelled with the best predicted mail folders for the current message. This increased the probability of offering a correct folder and also ensured the cost of an incorrect classification was minimal.

We have taken a somewhat similar approach. We sort each message in the *inbox* according to its predicted classification. Figure 1 shows an example of an **i-ems** screen. The long left panel lists the user's folders. To the right of this, there are two main parts to the screen. The upper part shows the messages in the folder that is currently selected. In the figure, this is the inbox. This shows each message under its predicted folder category or under the category *Unknown* if no prediction has be made. For example, the top three mail messages have been predicted to belong in the *Friends* folder. Each appears as a single line with the sender and the subject. The user has currently selected the single mail item classified under *Research*. This is displayed in the lower right-hand window.

Once a user has read a message, there are two possible courses of action. If they are happy with the classification, they can simply click on the *Archive* button at the top left of the screen. In the case of the current message shown in Figure 1, the *Archive* button would move it to the *Research* folder.

[1] Including the widely used programs Netscape and Microsoft Explorer.

IUI'02, January 13-16, 2002, San Francisco, California, USA.

The other possible case is that the user is not happy with the classification. Then the user selects the *MoveTo* button followed by the name of the folder in the left panel. This moves the message to the correct folder. This is much the same amount of effort required for a user to archive messages in a standard email manager.

This interface should reduce the cognitive drudgery and the time taken to archive each mail message. If the system makes the correct classification, the user simply accepts that with a single click. If the system is wrong, the user does the classification task they would have had to do anyway. This should significantly reduce the cost of incorrect classification while improving the overall interaction with the email manager.

Ducheneaut and Bellotti [2] found that many users tend to organise email into folders around one or more of the following criteria: Sender, Organisation (e.g a client or professional body), Project or Personal Interests. Learning to classify email into folders based on the Sender is simple, but the other types of classification are much more difficult. Hence the accuracy of automated classification is likely to differ significantly between folders.

Information regarding the accuracy of classification may also be provided to the user. **i-ems** provides the user with the percentage the folder has correctly classified in the past along with the percentage of messages the classifier has missed from that folder. This may be seen in Figure 1.

Another way in which **i-ems** allows the user to scrutinise the classifier is by showing the user the reason why an email was classified into a particular folder. When the user clicks on an email, it is displayed along with the reason for the classification. The explanation is displayed in the middle right hand panel(as seen in Figure 1).

This raises the important point of the scrutability of the explanations of different classifiers. We have implemented four different classifiers: Sender, Keyword, TF-IDF and DTree. Each provides some explaination for classification of individual messges. For example the explanation generated by the *TF-IDF* learner for the message shown in Figure 1 would be something like : *"Language", "ryan", and "we" are significate in predicting "Research".*

Users may also view the classifier to scrutinise it's classification. Figure 2 shows a classifier generated by the *Keyword* learner.

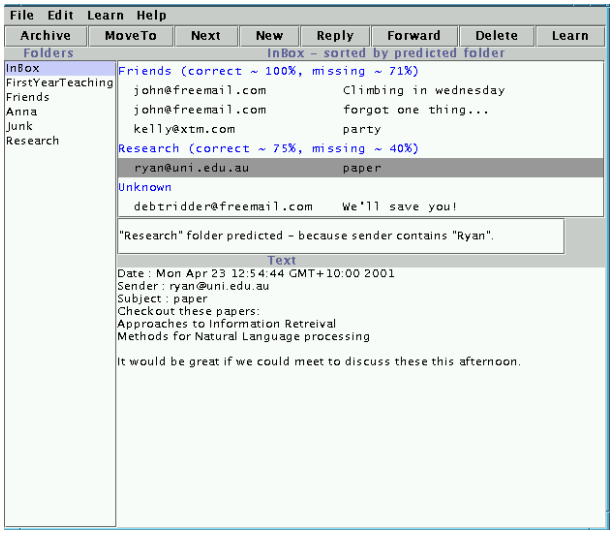

Figure 1: A screen shot of the i-ems interface.

```
filter(M,Anna) <= contains(M,sender,anna).
filter(M,Research) <= contains(M,sender,Ryan),contains(M,subject,Conference).
filter(M,Research) <= contains(M,sender,Ryan),contains(M,subject,paper).
filter(M,Research) <= contains(M,sender,jane).
filter(M,Research) <= contains(M,subject,code).
filter(M,Junk) <= contains(M,body,free).
filter(M,Junk) <= contains(M,body,today).
filter(M,FirstYearTeaching) <= contains(M,body,am).
filter(M,FirstYearTeaching) <= contains(M,body,Thanks),contains(M,sender,tammy).
filter(M,FirstYearTeaching) <= contains(M,sender,tbat).
filter(M,FirstYearTeaching) <= contains(M,sender,sam).
filter(M,Friends).
```

Figure 2: An example set of rules generated by i-ems using the keyword learning approach.

DISCUSSION AND CONCLUSIONS

It seems likely that there are some mail classifications which are relatively easy for an automated tool to learn. Empirical results suggest there are categories which can be learnt with high precision using very simple strategies[1].

We have also presented the **i-ems** interface which we are using to explore how to integrate automatic classification into a useful and usable interface. This is designed to take account of the limitations of automatic classification.

REFERENCES

1. Crawford, E. and Kay, J. and McCreath, E. Automatic Induction of Rules for e-mail Classification In *ADCS2001 Proceedings of the Sixth Australasian Document Computing Symposium*, pages 13–20, Coffs Harbour, NSW Australia, 2001

2. Ducheneaut, N. and V. Bellotti E-mail as habitat: an exploration of embedded personal information management In *Interactions*, pages 30–38, Volume 8, Number 5, 2001, ACM Press.

3. Mackay, W.E. Triggers and barriers to customizing software In *CHI'91 Conference on Human Factors in Computing Systems*, pages 153–160, New Orleans, Louisiana, 1991.

4. R.Segal and M. Kephart Mailcat: An intelligent assistant for organizing e-mail In *Proc. of The Third International Conference on Autonomous Agents*, pages 276-282, Seattle, WA, 1999.

5. S. Whittaker and C. Sidner. Email overload: Exploring personal information management of email. In *CHI*, pages 276-283, 1996.