The Intelligent E-Mail Sorter

Eric McCreath
Department of Computer Science
Faculty of Engineering and Information Technology
College of Engineering and Computer Science
The Australian National University
Australia

Introduction

- Email has become our habitat.
  - Ducheneaut/Bellotti 2001 ACM interactions
- Web based approaches are very common.
  - Hotmail - 200 million active accounts - 75 million emails a day. (http://advertising.msn.co.uk)
- A large number of different uses:
  - conversation, task manager, document delivery, contact manager....
- Email is accessible, universal, and robust....
- Use of instant messaging is growing.
- In all this we see a converging of technologies.

People/History

- 6 conference papers and 1 journal article.
- Started with collaboration with Judy Kay.
- Research grant $20K.
- Employ Liz Crawford (currently doing a PhD at CMU)
Overview

- User Interface
- Learning Approach
- Evaluation Metric
- The new email manager.
- Some new directions:
  - Learning importance.
  - Combining email with a Todo application.
  - Organizing different types of information using the same categories.
  - Exploring different user interfaces.

User Interface

- Different people organize their email in very different ways.
- Q: Should the interface move messages into the folders? A: YES.
- Q: Should the interface leave the messages in the inbox? A: YES
- The interface should be:
  - Scrutable,
  - Modifiable,
  - Efficient, and
  - Predictable.

Experimental Results

- Comparing handcrafted and learnt rules.
- Single user over 3 months, 5100 messages, 70 hand crafted messages, 21 different folders.
Learning Approach

- Many learners form explicit generalization of the training data.

Instance based learning approaches simply collect the training examples. No explicit generalization is constructed from the data. Rather when a prediction is needed on a new instance, the distance from the instance to all of the training examples is measured.

Our approach generates an explicit hypothesis that directs the use of an instance based approach.

These rules are augmented with hand crafted rules:

- \( \text{filter}(M,F) \leftarrow \text{lastXsender placements}(F,M,5), F \neq \text{“ml”} \).
- \( \text{filter}(M,F) \leftarrow \text{contains_same}(F,M,\text{subject}), F \neq \text{“colleagues”} \).

More Experimental Results

- 5 users
- Six approaches compared:
  - Keyword,
  - TF-IDF,
  - Dtree,
  - Naive Bayes, and
  - Composite Rule.

<table>
<thead>
<tr>
<th>User</th>
<th># Messages</th>
<th># Folders</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>526</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>945</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>682</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>429</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>15000</td>
<td>24</td>
</tr>
</tbody>
</table>
Measuring Success

- How do you measure the success of such a system?
  - accuracy,
  - precision,
  - recall,
  - f1, or
  - other ???

The metric used to evaluate the performance of a system is also reflected back into how the system selects competing hypotheses.

Experimental Results

Percentage of Unknown Classifications

<table>
<thead>
<tr>
<th>User</th>
<th>Sender</th>
<th>KeyWord</th>
<th>TF-IDF</th>
<th>Dtree</th>
<th>Bayes</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.5</td>
<td>69.4</td>
<td>62.7</td>
<td>68.2</td>
<td>56.3</td>
<td>64.7</td>
</tr>
<tr>
<td>2</td>
<td>51.4</td>
<td>43.7</td>
<td>46.5</td>
<td>61.4</td>
<td>38.3</td>
<td>49.5</td>
</tr>
<tr>
<td>3</td>
<td>28.5</td>
<td>41.3</td>
<td>29.2</td>
<td>48.8</td>
<td>19.7</td>
<td>23.8</td>
</tr>
<tr>
<td>4</td>
<td>46.3</td>
<td>49.0</td>
<td>53.9</td>
<td>-</td>
<td>12.2</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>62.4</td>
<td>67.1</td>
<td>62.7</td>
<td>68.2</td>
<td>65.5</td>
<td>65.5</td>
</tr>
<tr>
<td>6</td>
<td>36.7</td>
<td>42.4</td>
<td>29.2</td>
<td>46.9</td>
<td>27.1</td>
<td>4.5</td>
</tr>
<tr>
<td>7</td>
<td>55.5</td>
<td>58.2</td>
<td>53.9</td>
<td>-</td>
<td>18.5</td>
<td>18.5</td>
</tr>
</tbody>
</table>

Recall

<table>
<thead>
<tr>
<th>User</th>
<th>Sender</th>
<th>KeyWord</th>
<th>TF-IDF</th>
<th>Dtree</th>
<th>Bayes</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.5</td>
<td>69.4</td>
<td>62.7</td>
<td>68.2</td>
<td>56.3</td>
<td>64.7</td>
</tr>
<tr>
<td>2</td>
<td>51.4</td>
<td>43.7</td>
<td>46.5</td>
<td>61.4</td>
<td>38.3</td>
<td>49.5</td>
</tr>
<tr>
<td>3</td>
<td>28.5</td>
<td>41.3</td>
<td>29.2</td>
<td>48.8</td>
<td>19.7</td>
<td>23.8</td>
</tr>
<tr>
<td>4</td>
<td>46.3</td>
<td>49.0</td>
<td>53.9</td>
<td>-</td>
<td>12.2</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>62.4</td>
<td>67.1</td>
<td>62.7</td>
<td>68.2</td>
<td>65.5</td>
<td>65.5</td>
</tr>
</tbody>
</table>

Precision

<table>
<thead>
<tr>
<th>User</th>
<th>Sender</th>
<th>KeyWord</th>
<th>TF-IDF</th>
<th>Dtree</th>
<th>Bayes</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.5</td>
<td>69.4</td>
<td>62.7</td>
<td>68.2</td>
<td>56.3</td>
<td>64.7</td>
</tr>
<tr>
<td>2</td>
<td>51.4</td>
<td>43.7</td>
<td>46.5</td>
<td>61.4</td>
<td>38.3</td>
<td>49.5</td>
</tr>
<tr>
<td>3</td>
<td>28.5</td>
<td>41.3</td>
<td>29.2</td>
<td>48.8</td>
<td>19.7</td>
<td>23.8</td>
</tr>
<tr>
<td>4</td>
<td>46.3</td>
<td>49.0</td>
<td>53.9</td>
<td>-</td>
<td>12.2</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>62.4</td>
<td>67.1</td>
<td>62.7</td>
<td>68.2</td>
<td>65.5</td>
<td>65.5</td>
</tr>
</tbody>
</table>

F1 Measure (Harmonic mean of recall and precision)

<table>
<thead>
<tr>
<th>User</th>
<th>Sender</th>
<th>KeyWord</th>
<th>TF-IDF</th>
<th>Dtree</th>
<th>Bayes</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.5</td>
<td>69.4</td>
<td>62.7</td>
<td>68.2</td>
<td>56.3</td>
<td>64.7</td>
</tr>
<tr>
<td>2</td>
<td>51.4</td>
<td>43.7</td>
<td>46.5</td>
<td>61.4</td>
<td>38.3</td>
<td>49.5</td>
</tr>
<tr>
<td>3</td>
<td>28.5</td>
<td>41.3</td>
<td>29.2</td>
<td>48.8</td>
<td>19.7</td>
<td>23.8</td>
</tr>
<tr>
<td>4</td>
<td>46.3</td>
<td>49.0</td>
<td>53.9</td>
<td>-</td>
<td>12.2</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>62.4</td>
<td>67.1</td>
<td>62.7</td>
<td>68.2</td>
<td>65.5</td>
<td>65.5</td>
</tr>
</tbody>
</table>
Future Direction

- Process of writing a completely new version of IEMS. (Called SE-MAIL or MyMail or ??)
- Managing tasks using email.
- Learning the importance of incoming messages.
- Moving away from the bag of words representation for text.

The New Email Manager

- Uses IMAP as a backing store for messages and meta data.
- Written in Java uses the JavaMail API
- 51 Class, ~5000 lines of code and growing!
- Uses both hand crafted and learnt rules.(basic sender learner only implemented.)
The Todo List

People use the flagging of email to manage their work, however, this approach is often limited.

A Unified View of Desktop Artifacts

- Categorization of your desktop artifacts provides a way of organizing your work.
- These categories can go across different types of desktop artifacts including: email, todos, notes, addresses, and even files.
- By unifying interface in which a user interacts with these 'blobs' of information it is hoped one can reduce:
  - the amount of interaction required to achieve tasks, and
  - the cognitive load on the user.

Files

- We have lots of files. Lots and lots of files. It can be often hard to track down the file you are looking for. (I have 115,245 files, 25 files a day over the last 13 years)
- Often emails stores act as a users file system. Why not merge them and do away with our traditional view on files?
- Consider the following scenario:

  You are sent an email with an attached open office document that you need to modify and send back.

Conclusion

- Change is on the way.