Intelligent Wireless Information Access

Haym Hirsh
Computer Science Department
Rutgers University
www.cs.rutgers.edu/~hirsh
Outline

• Information Valets ("iValets")
• The Rutgers iValet
• Learning Users’ Wireless Preferences

See also my talk “Integrating Multiple Sources of Information in Text Classification” tomorrow at 11:30am in 101 Warren Weaver Hall
Premises

• Computers as tools for information access
• Ubiquity of wireless platforms
  – Cellphones
  – Pagers
  – PDAs (Palm VII, etc.)
  – Laptops with wireless modems
  – ….
• Multi-device access to information
Device Heterogeneity

Information heterogeneity is understood
Device heterogeneity is also very important
• Device connectivity characteristics
• Device capabilities
• Device state
Device Connectivity Characteristics

• How is the device connected to its server
  – Always on/“pushable”?  
  – Pagers, cellphones vs Palm VII
• Bandwidth
• Cost
Device Capabilities

What can the client device do?

• “Display” characteristics
  – Text, audio, images, video, etc.
  – Screen size, resolution, etc.

• Local storage? Maximum item size?

• Processor speed

• Ports? (Infrared? BlueTooth?)
Device State

- Battery strength
- Remaining storage
- Connected?
- Price per byte
Information Valets (iValets)

A framework for intelligent information access to information from a range of platforms for a range of information services

iValet: the central conduit for accessing and manipulating information:

– iValet interacts with user via multiple client devices
– iValet interacts with multiple information resources

Understands heterogeneity of information and devices
iValet Framework

Client Devices

- Desktop PC
- Connected Organizer
- Two-way Pager
- Cellphone

Information Services

- Email
- Web Content
- Documents
- Printing Faxing

Etc.

Etc.
Design Goals

• Incremental addition of devices
• Incremental addition of information services
iValet Architecture
The Rutgers iValet

- User devices:
  - RIM 950 two-way pagers
  - Palm VII

- Information services
  - Email
  - Web
  - Files
Front Ends

RIM 950
- BellSouth Wireless Data Inter@ctive Paging Service
- Communicates with iValet via email
- Uses features of email standard:
  - Reply-to field
  - user+action@machine, user+id@machine
- In use 18 months

Palm VII
- Web-based front-page
- In use 6 months
Back Ends

- Reading, writing, and searching email
- Searching files by name and content
- Text-based URL browsing
- Printing
Rutgers iValet Core

- Allows cross-device functionality
- Location where *intelligent* wireless information access happens
User Information

Intelligent information access requires information about the user

• User context (e.g., location, time)
• User model (learned vs provided, content-based vs collaborative, device-specific vs device-independent, short-term vs long-term)
• User feedback channels
Learning Users’ Wireless Information Access Preferences

• Can we predict what to do with new incoming information?
• What kinds of user feedback can be helpful?
• What context helps with learning user preferences?
• What learners are well-suited to this task?
Experimental Set-Up

• Record which messages user read on RIM 950
• Predict whether each message would have been read by learning from preceding email
• Evaluation: 500 random messages for each user, learn from preceding email
Learning Methods

• Rainbow: IR methods
  – Naïve Bayes
  – Maximum Entropy
  – TF-IDF
  – Pr TF-IDF

• Ripper: Rule learning

• Winnow: Linear threshold

• Remove null-body messages

• Stemming and Stoplisting
Data

- Three users (the authors)
- 6+ Months of use
- 25,000+ messages total

<table>
<thead>
<tr>
<th>User</th>
<th># Messages</th>
<th>% forwarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aynur</td>
<td>2856 (1499)</td>
<td>26.2% (26.4%)</td>
</tr>
<tr>
<td>Haym</td>
<td>7406 (5322)</td>
<td>23.9% (26.3%)</td>
</tr>
<tr>
<td>Sofus</td>
<td>2370 (779)</td>
<td>14.8% (15.5%)</td>
</tr>
</tbody>
</table>
Features Used

• Email Headers
  – Date, To, From, Subject, To&CC, Length

• Email Body

• Email Context
  – Time since sent email to person (on subject)
  – Time since received email from person (on subject)
  – Time since last pager command

• Online Context
  – Time since last online command
  – Time since last read mail online
  – Time since last logged in
Evaluation Criterion

- **Precision/Recall curves**
  - Precision: % messages forwarded that should be forwarded
  - Recall: % messages that should have been forwarded that were forwarded

- **Break-Even points as one objective measure**
  - Break-Even point: Where Precision and Recall are closest to being equal.
Experiments

• Experiment 1: Does any learning method work best?
• Experiment 2: What context features are important?
• Experiment 3: Window size effects
Summary

• Information Valets as a framework for multi-platform intelligent wireless information access

• Learning has the potential to model users’ information-access preferences
Future Work

• More comprehensive study of features
• Prioritizing versus pruning
• Security
• Putting it all together