Case Study for Supporting Intelligibility in Context-Aware Systems: IM Auto-Status Explanations

Verantw. Hochschullehrer: Prof. Dr. Heinrich Hußmann
Betreuerin: Sara Streng
Externer Betreuer: Prof. Anind Dey, Ph.D. (CMU)

30. Juni 2009
Outline

- Motivation
- Goals of this Work
- Realization of the Software
- User Study
- Conclusion and Outlook
Instant Messaging

- Important role in work environments
- Short interaction
  - Quick questions and clarifications
  - Coordination and scheduling
  - Impromptu social meetings
  - Contact with friends and family
- Longer interaction
  - Discussion of complex topics
  - Collaboration

- But also:
  - Receiver is disrupted by incoming messages
  - Sender has to wait for response
  - Negative effect on
    - Task performance
    - Memory
    - User’s emotional state
  - Cause of annoyance
Possible Solutions

For the receiver

- Analysis of incoming messages
- Notification according to importance
- Reduction of unwanted work interruptions
- Example: QnA

For the sender

- Consideration of receiver’s context
- Information about the receiver’s presence and availability
- Reduction of waiting times due to unresponsiveness
- Example: MyVine

[1] [6]
Intelligent Systems and Intelligibility

- Intelligent systems can never be perfect
- User acceptance depends on trust in the system
- Credibility can easily be lost!
  - When system does not behave as the user expects
  - Small errors can have relatively big impact
- Principles to support intelligibility
  - Information about system’s understandings
  - Feedback (feedforward and confirmation)
  - Identity and action disclosure
  - Control
- Promising approach: Increasing intelligibility by explanations
  - Used in product recommendation system (Pu & Chen)
Outline

• Motivation
• Goals of this Work
• Realization of the Software
• User Study
• Conclusion and Outlook
Goals of this Work

• „How can a context-aware system help the user to understand how it works?“
• Design and implementation of a software that
  • provides responsiveness information about IM users
  • is able to explain its predictions
• Design of a long-term study which
  • allows to evaluate different versions of explanation provision
  • can provide findings about understanding and usefulness
Outline

- Motivation
- Goals of this Work

**Realization of the Software**

- User Study
- Conclusion and Outlook
IM Auto-Status Plugin

- Plugin for AOL Instant Messenger (AIM)
- „Auto-Status“ capability
- Principle of work:
  - A sends B a message
  - B’s plugin predicts when B is likely to respond
  - The prediction is sent back to A
- User interface
  - Makes use of messaging window
  - No additional UI elements
Explanation feature

- 5 reasons for why the system predicted like it did
- Additional features:
  - Prediction certainty
  - Exact values for reasons (will be explained later)
- Interaction using textual commands
- Three different versions of the plugin showing
  - Explanations for each prediction
  - Explanations on demand
  - No explanations
Intelligibility in Context-Aware Systems: IM Auto-Status Explanations

Andreas Möller
moellera@ifi.lmu.de

Prediction

I will respond between 1 and 5 minutes.

Prediction with optional explanation

I will respond between 1 and 5 minutes.
You can type im-why, im-% or im-help.

Prediction with explanation in form of reasons (boundary values)

I will respond between 1 and 5 minutes because
1. I have more than 1 open window
2. ...
3. ...
4. ...
5. ...
You can type im-1, ..., im-5, im-% or im-help.

Level of detail

Actual (exact) values of each feature in the explanation

I have 3 open IM windows.

It is 80% certain that I will respond between 1 and 5 minutes.
You can type im-why or im-help.

Start level of detail in different versions

<table>
<thead>
<tr>
<th>No explanations</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanations on demand</td>
<td>2</td>
</tr>
<tr>
<td>Always explanations</td>
<td>3</td>
</tr>
</tbody>
</table>
How does it work?

- Permanent activity monitoring
  - Desktop events: typing, switching between applications, ...
  - IM-related: incoming messages, time until IMs are answered, ...

- Responsiveness prediction using machine learning
  - Snapshot creation ("instance") using 68 features
  - Classification of user's current state using a decision tree
  - Pre-trained model

- Explanation generation
  - Extraction of the 5 most important features
  - Transformation in prose English
Software Architecture: Schematic Overview

Intelligibility in Context-Aware Systems: IM Auto-Status Explanations
Andreas Möller
moellera@ifi.lmu.de

Activity Monitor
Classifier
Explanation Generator
Messaging API
Survey Generator

Logging Subsystem
Hard Disk
Web Server

User Level
Chat Partner
Plugin User
Incoming message
Automated prediction sending
Command interaction
Regular automated surveys

System Level
C#
Java
Activity Monitor
Classifier
Explanation Generator
Logging Subsystem

Continuous activity & behavior sensing
Classification request
Responsiveness prediction

Current state snapshot
Local log files
SFTP uploads

Web Server
Outline

- Motivation
- Goals of this Work
- Realization of the Software
- User Study
- Conclusion and Outlook
Hypotheses

- Recap: Three versions of explanation provision:
  - Always
  - On demand
  - Not at all (predictions only)

- **H1**: Users who get explanations find the application more useful and are more satisfied when using it.

- **H2**: Compared to users of the on-demand version, users who always get explanations will understand the application better. Yet, they will find it more obtrusive than users of the on-demand version.

- **H3**: Subjects using the low accuracy version find the explanations more useful, and improve more than those using the high accuracy version.
Study Setup

- 12 groups, each consisting of at least 3 participants
- One person in each group installs the plugin (receiver), the others receive predictions about him/her (senders)
- Daily chats during 2 phases of 3 weeks each
- Two different explanation types in phase 1 and 2

<table>
<thead>
<tr>
<th>Group</th>
<th>Prediction Accuracy</th>
<th>Explanation type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Phase 1</td>
</tr>
<tr>
<td>1</td>
<td>High</td>
<td>No explanations</td>
</tr>
<tr>
<td>2</td>
<td>high</td>
<td>No explanations</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>No explanations</td>
</tr>
<tr>
<td>4</td>
<td>Always On</td>
<td>No explanations</td>
</tr>
<tr>
<td>5</td>
<td>Always On</td>
<td>No explanations</td>
</tr>
<tr>
<td>6</td>
<td>On Demand</td>
<td>No explanations</td>
</tr>
<tr>
<td>7</td>
<td>Low</td>
<td>No explanations</td>
</tr>
<tr>
<td>8</td>
<td>Low</td>
<td>No explanations</td>
</tr>
<tr>
<td>9</td>
<td>Low</td>
<td>No explanations</td>
</tr>
<tr>
<td>10</td>
<td>Always On</td>
<td>No explanations</td>
</tr>
<tr>
<td>11</td>
<td>Always On</td>
<td>No explanations</td>
</tr>
<tr>
<td>12</td>
<td>On Demand</td>
<td>No explanations</td>
</tr>
</tbody>
</table>
Data Collection

- Regular surveys
  - Daily questions, referring to particular IMs
    - Agreement with prediction and user predictions
    - User understanding
  - Weekly online questionnaires, referring to the overall impression
    - User understanding
    - Perception of usefulness, reliability, and trust
    - Changes in interaction with buddies (qualitative)
- Personal interview (phone or IM)
- Logging by the plugin
  - Actual correctness of the predictions
  - Timing/behavior pattern changes
The study is not yet completed.

- **Why?**
  - Recruitment
    - Less interested parties than expected → extension to other cities
  - Subscription phase
    - Incomplete/bogus subscriptions
    - Groups not meeting all requirements
  - Decline of agreement terms („daily chats“)
  - Technical issues
- **Difficulties in extensive long-term studies like this:**
  - High effort for participants
  - Group subscriptions
  - Reduced commitment due to missing face-to-face relationship
First Trends (1)

- Usefulness of predictions and explanations

1: strongly disagree, 5: strongly agree

I find the predictions useful.

<table>
<thead>
<tr>
<th>None</th>
<th>On Demand</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.79</td>
<td>3.20</td>
<td>3.00</td>
</tr>
</tbody>
</table>

I find the explanations useful.

<table>
<thead>
<tr>
<th>On Demand</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.30</td>
<td>2.89</td>
</tr>
</tbody>
</table>

1: strongly disagree, 5: strongly agree
First Trends (2)

• Preference for respective other version

Would you rather have always explanations?

- Yes: 40%
- No: 40%
- Not Sure: 20%

Would you rather have explanations on demand?

- Yes: 89%
- No: 11%
- Not Sure: 0%

among subjects who get explanations on demand

among subjects who get always explanations
First Trends: Summary

- Predictions are appreciated
- Usefulness increases with explanations
- Strong preference towards on-demand version
  - Usefulness of explanations considered higher
  - „Always explanations“ too obtrusive
  - No advantages for „always explanations“ shown (yet)
- H1 and H2 are likely to be verified
Outline

- Motivation
- Goals of this Work
- Realization of the Software
- User Study
- Conclusion and Outlook
Conclusion and Outlook

• Responsiveness predictions and explanations
  • Can help senders to estimate response time
  • Can help reduce work interruptions

• Future of this project
  • Relation between effect of explanations & error rate
  • Changes in understanding and behavior

• Outlook
  • Downloadable version at AIM Gallery?
  • Different approach: change salience of message notifications according to responsiveness
Thank you for your attention!

Questions and Discussion
References


