WhoAmI: Profiling Attendees before Meetings

Jianqiang Shen, Oliver Brdiczka, Juan Liu, Masafumi Suzuki
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304, USA
{jianqiang.shen, oliver.brdiczka, juan.liu, masafumi.suzuki}@parc.com

ABSTRACT

Mobile phones have been widely used to access, retrieve and organize information. In this paper, we present the *meshin* system, an intelligent personal assistant that organizes messages, notifications and appointments for mobile phone users. To help users prepare for meetings, meshin automatically creates a profile for each meeting attendee. The meshin system searches the Internet with the attendee's name and the domain of the email address, then retrieves and aggregates information for this attendee. To get a better understanding of the attendee, it further estimates the personality profile based on the emails he/she wrote. Our experimental results show that our system can predict personality with reasonable accuracies (95%).

Author Keywords

Personal assistant; email; personality; text processing

ACM Classification Keywords

J.4 Social and Behavioral Sciences: Psychology

## General Terms

Human Factors, Experimentation

INTRODUCTION

Knowledge workers, such as professors, managers, and engineers, perform many different activities in their daily life, and are struggling with information overload. They spend a significant amount of time processing electronic messages and other notifications - in fact, previous research even characterized email as the natural "habitat" of the modern office worker [1]. The growing importance and popularity of electronic communication has inspired many attempts to develop intelligent tools for presenting, organizing, classifying, and prioritizing electronic messages. Such systems help multi-tasking knowledge workers easily organize and access messages and other resources (contacts, files, web pages, etc.) that they need to support their activities.
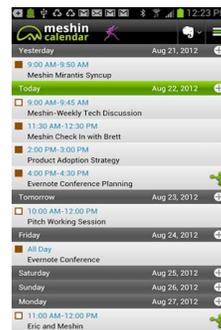
Meshin is a personal assistant developed for Android cell phones that hopes to "bring order to communication chaos".
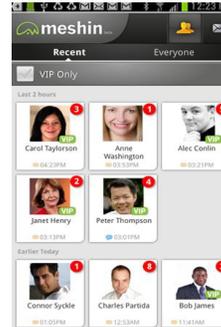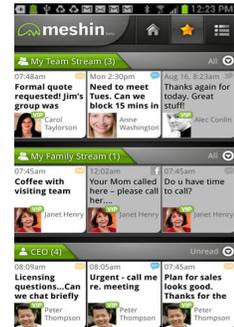
(a) VIP List (b) Message Stream

(c) meshin Calendar (d) Event Detail

Figure 1. Screenshots of meshin, a mobile personal assistant application.

It aggregates various information sources into one easy-to-navigate screen and acts as a communication hub that unifies emails, phone calls, text messages, social network streams, calendar events and notes into a centralized application. The meshin system is trying to tackle the growing challenge of information overload by separating the important information from the "noise" in several ways:

- **VIP List**. meshin learns which contacts are more important, so that their messages and interactions appear higher in the list of communications. It also allows you to set VIPs for certain people manually. See Figure 1(a).

- **Message Stream** aggregates your messages from various applications into one neat view. It allows you to sort these various messages by people or source. There's no need to switch back and forth from different messaging applications. See Figure 1(b).

- **The meshin Calendar** allows you to view multiple calendars at once and links your calendars with Evernote so you can make text, photo, and audio notes, attach them to appointments, and save them to your Evernote account for quick reference. See Figure 1(c).
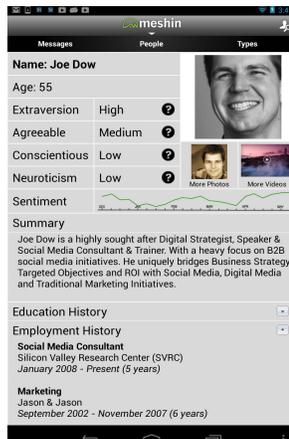
Figure 2. WhoAmI: meeting attendee profile generated by meshin.

- **Event Detail**. Meshin analyzes the meeting agenda, attendees and location to automatically provide prompts for missing logistical information, while suggesting contextually relevant information from Evernote. See Figure 1(d).

One major design goal of meshin is to help users with notifications and messages. Email is widely used to schedule meetings and appointments. It is advantageous to know more about attendees beforehand, so users often search for attendee information online before the meeting. However, web browsing using mobile phones is time-consuming. Further, accomplishing information tasks with mobile phones can be difficult due to screen size limitations. A user, in particular when mobile, would prefer to extract the desired information quickly and with minimal mental effort [3]. The meshin system thus automatically creates a profile for each meeting attendee.

MEETING ATTENDEE PROFILING
Figure 2 shows the WhoAmI UI of the meeting attendee profiling that meshin generates. It smartly searches the Internet with the attendee's name and the email domain, retrieves possible biographies, photos, videos, and Twitter/Facebook posts, to help the user know more background about this attendee. From the email messages and posts that have been written by the attendee, meshin calculates positive/negative words and analyzes the attendee's sentiment fluctuation, which could reflect this attendee's emotional status. Further, we estimate the attendee's personality based on the emails he/she wrote.

Personality traits are consistent patterns of thoughts, feelings, or actions that distinguish people from one another [2]. We posit that some personalities are more likely to react with negative actions to job and life stress than others. Consequently, we introduce personality traits as part of the attendee profile for the user to review before a meeting. This could enable the user to identify users for good team building or avoid potential personality conflicts beforehand. The Big 5 personality model believes that there are five factors accounting for most individual differences in personality [2]. In this paper, we focus on predicting four components of the Big 5 model including Neuroticism, Agreeableness, Conscientiousness and Extraversion from the content of written emails:

- **Neuroticism** relates to emotional stability.

- **Agreeableness** is a tendency to be compassionate and cooperative rather than suspicious and antagonistic.
- **Conscientiousness** describes socially prescribed impulse control that facilitates task and goal-directed behavior.
- **Extraversion** implies an energetic approach to the social world and includes traits such as sociability.

For each email sent from the attendee, we first perform some clean-up process before extracting its content. We detect *reply lines* and *signature blocks* based on both "false positive" regular expressions and pre-trained automatic reply/signature detectors. We then calculate meta features about the message, including count of TO/CC/BCC recipients, count of different punctuation symbols, count of words, count of characters, count of positive and negative numbers, count of pronouns, count of negations. If it is a reply email, we then calculated the duration since he/she received the original email, and those meta features in the original email.

We created a *common word list* with the most common words from TV and movie scripts based on Wiktionary by removing names, addresses, and job titles. This common word list contains 16,623 words. Instead of collecting every appearing word, we only build our features from words that are most commonly used in everyday life. Research on sentiment analysis has shown that common words are most predictive to identify the viewpoints underlying a text span. Focusing on common words can also help to avoid associating specific, unique words to certain individuals in the training set and not to certain personalities. We apply *part-of-speech* (POS) tagging to the cleaned email and record the count of words for each POS tag. We adopt *sentiment analysis* to count the number of words in different sentiment categories. We analyze the writing style of the email, such as whether containing an open greeting, count of smileys. Additionally, we use a pre-trained predictor to predict Speech Acts from the email and adopt the prediction confidence of each act as a feature.

For each personality trait, we train an SVM predictor using the above features. This predictor estimates a probability distribution on trait values based on one single message. We then use those estimated probabilities to vote for the final prediction. We rank trait values based on their overall scores and make predictions with those having highest scores. The predictor was trained using a large email data set collected by us, containing 458 valid users and 65,185 emails. We evaluated the prediction accuracy based on 10-fold cross validation with the constraint that emails from the same user either went to the training set or the test set. When using one message for prediction, its average accuracy for predicting the 4 personality traits is 72%. The accuracy is nearly 90% if we see more than 10 messages, and 95% if we see more than 35 messages.

REFERENCES
1. Ducheneaut, N., and Bellotti, V. E-mail as habitat: an exploration of embedded personal information management. *Interactions 8* (2001), 30–38.
2. John, O. P., Robins, R. W., and Pervin, L. A. *Handbook of Personality: Theory and Research*. The Guilford Press, 2010.
3. Setlur, V., Rossoff, S., and Gooch, B. Wish i hadn't clicked that: context based icons for mobile web navigation and directed search tasks. In *Proc. of the 16th IUI* (2011), 165–174.