

Vorhersagbarkeit von Selektiver Verfügbarkeit im Instant Messaging

Mirko Fetter, Julian Seifert, Tom Gross

Bauhaus-Universität Weimar

Zusammenfassung

Gegenwärtige Instant Messaging-Systeme stellen NutzerInnen beim Einstellen ihrer Verfügbarkeit vor zwei Herausforderungen. Zum einen erlaubt ein globaler Onlinestatus nur pauschale Darstellung der eigenen Verfügbarkeit gegenüber unterschiedlichen sozialen Kontakten. Zum anderen muss der Onlinestatus kontinuierlich manuell der tatsächlichen Verfügbarkeit angepasst werden. In dieser Arbeit zeigen wir mittels einer vierwöchigen Experience Sampling Studie, dass ein Bedarf nach Ausdifferenzierung besteht und wie die Adaptation selektiver Onlinestatus automatisiert werden kann.

1 Einleitung

Der zunehmende Einzug von computergestützter Kommunikationstechnologie in unseren Alltag erlaubt es schneller und einfacher zu kommunizieren. Allerdings bringen diese Technologien, welche ständige Verfügbarkeit ermöglichen, dadurch auch Einschränkungen mit sich. Der oft subtile Aushandlungsprozess, der in der direkten Interaktion Menschen den adäquaten Moment für eine Aufnahme von Kommunikation ermitteln lässt, wird in computergestützter Kommunikation oft unzureichend unterstützt. Gerade bei synchronen Kommunikationstechnologien ist das Resultat dabei nicht selten die Unterbrechung des Kontaktierten in einer Tätigkeit oder unpassenden Situation und somit das Scheitern der Kommunikation (Volda et al. 2002).

Im Bereich der Instant Messaging (IM) Systeme unterstützt der Onlinestatus (z.B.: Away, Online, Do Not Disturb) sowie die Statusmeldung den Kontaktaufnehmenden darin, einen geeigneten Zeitpunkt für die Kommunikation besser abzuschätzen.

Gegenwärtige IM-Systeme stellen die Nutzer dabei vor zwei Herausforderungen:

- Aktuelle IM-Systeme bieten für einen Benutzerzugang nur einen pauschalen Onlinestatus an. Die Nutzer sind folglich jederzeit für alle Kontakte im gleichen Maße erreichbar. Eine Ausdifferenzierung der Verfügbarkeit wie beispielsweise nach Gruppen mit unterschiedlicher sozialer Bindung (Familie, Arbeitskollegen, etc.), ist in momentanen Systemen nur

über Umwege möglich. So zeigen zum Beispiel Volda et al. (2002) in ihrer Studie, dass IM-Nutzer durch den Einsatz mehrerer Benutzerzugänge ihre Kontakte in soziale Gruppen, aufteilen um so ihre Verfügbarkeit selektiv zu verwalten. Ähnliche Erkenntnisse liefern auch andere Studien (Davis & Gutwin 2005; Patil & Kobsa 2004). Aktuelle Forschungsarbeiten zeigen dafür erste Lösungsansätze (Fetter et al. 2008).

- Der Onlinestatus muss von den BenutzerInnen per Hand eingestellt werden, was in einer Unterbrechung des Arbeitsflusses resultiert und oft vergessen wird. Das Ergebnis ist ein unzutreffender Onlinestatus, der in Folge dessen für die Kontaktaufnehmenden seine Aussagekraft verliert und somit ignoriert wird. Dieses Verhalten wurde unter anderem in einer Studie aufgezeigt, die darlegt wie Nutzer durch Nachfragen die tatsächliche Verfügbarkeit erst in Erfahrung bringen (Volda et al. 2002). Milewski und Smith (2000) stellen in einer Benutzerstudie eines mobilen 'Personal Presence' Systems fest, dass die Teilnehmer ihren Status im Durchschnitt nur 1,4 mal am Tag aktualisieren. Einfache Heuristiken wie automatisierte Statusanpassungen wenn die BenutzerInnen eine gewissen Zeit nicht am Rechner gearbeitet haben, bieten erste Verbesserungen.

In dieser Arbeit präsentieren wir Ergebnisse, welche zeigen, wie es BenutzerInnen ermöglicht wird, die Vorteile von selektiver Verfügbarkeit zu nutzen und dabei gleichzeitig den Konfigurationsaufwand für die Einstellung der Verfügbarkeit zu reduzieren. Dabei beschreiben wir zunächst, wie die Daten für die Untersuchung gesammelt sowie für maschinelles Lernen vorbereitet wurden. Wir zeigen anhand der Ergebnisse einer vierwöchigen Experience Sampling Studie, dass ein Bedarf nach selektiver Verfügbarkeit besteht und wie durch die Nutzung von Verfahren des maschinellen Lernens die Adaptation der Onlinestatus automatisiert werden kann. Abschließend geben wir einen Überblick über verwandte Arbeiten und eine Zusammenfassung.

2 Experience Sampling für maschinelles Lernen

Um die Vorhersagbarkeit von selektiver Verfügbarkeit zu evaluieren, bedarf es zunächst eines geeigneten Trainingsdatensatzes. Dieser Datensatz wurde von uns in einer vierwöchigen Studie erhoben—durch den Einsatz der Experience Sampling Methode und der gleichzeitigen Erfassung von Sensordaten—und für das maschinelle Lernen aufbereitet.

2.1 Experience Sampling Setup

Die in der Sozialpsychologie entwickelte Experience Sampling Methode (ESM) (Csikszentmihalyi & Larson 1987) erlaubt es, Erlebnisse oder innere Zustände von Teilnehmern an einer Studie zum Entstehungszeitpunkt festzuhalten. In der Mensch-Computer-Interaktion wurde dieser Ansatz unterschiedlich aufgegriffen. Beispielsweise zur Erfassung der inneren Zustände von NutzerInnen in Kombination mit der Aufnahme von Sensordaten, um Trainingsdaten für die Entwicklung maschineller Lernverfahren zu erhalten (Horvitz et al. 2004; Kapoor & Horvitz 2007). Dieser Ansatz bietet die Grundlage unserer Studie, in der wir basierend auf ESM wiederkehrende Fragen zur Benutzerelbsteinschätzung hinsichtlich

der eigenen selektiven Verfügbarkeit stellen und über die Dauer der Studie die Daten von 34 Sensoren auf dem mobilen Rechner der Teilnehmer kontinuierlich mitloggen.

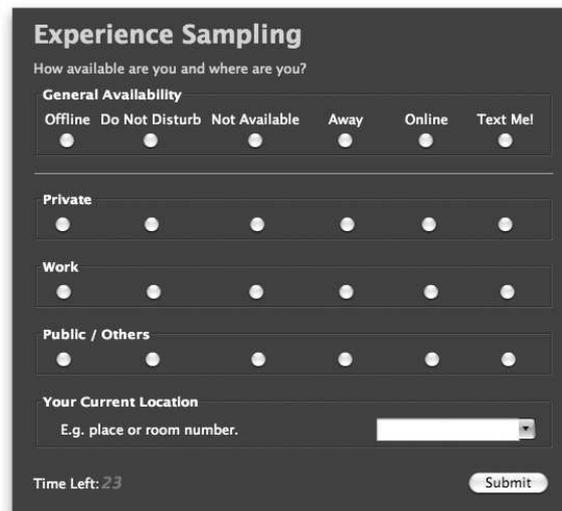


Abbildung 1. Dialogfenster des Experience Sampling.

Das Experience Sampling geschieht in Form eines Dialogfensters (vgl. Abbildung 1) das alle 25-35 Minuten die ProbandInnen auffordert, Angaben zu ihrer aktuellen Verfügbarkeit im Allgemeinen und im Speziellen hinsichtlich dreier Verfügbarkeitskategorien zu machen. Verfügbarkeitskategorien sind dabei eine Menge von Kontakten, für welche die BenutzerInnen im gleichen Maße verfügbar sind. Um die Vergleichbarkeit unter den ProbandInnen zu gewährleisten, wurden drei Kategorien, die in einer Vorstudie ermittelt wurden, verwendet: *Private*, *Work*, *Public/Others*. Mittels sechs Optionsfelder können die ProbandInnen jeweils Angaben über ihren aktuellen Verfügbarkeitslevel mit den Ausprägungen *Offline*, *Do Not Disturb*, *Not Available*, *Away*, *Online* und *Text Me!* machen und in einem Textfeld den aktuellen Aufenthaltsort eingeben.

2.2 Eingesetzte Sensoren

Ein Satz von 34 Sensoren nimmt über die gesamte Studiendauer kontinuierlich Daten der mobilen Rechnern in Form von Sensorevents auf. Die Frequenz jedes Sensors wurde dabei hinsichtlich des Ressourcenverbrauchs und der Aussagekraft ausbalanciert und befindet sich im Bereich von 30 bis 150 Sekunden. Ausgangspunkt für die Auswahl der verwendeten Sensoren ist eine Literaturrecherche (u.a. (Fogarty et al. 2005; Horvitz et al. 2004)). Die 34 Sensoren erfassen das Folgende:

- Aktuell am Rechner angeschlossen Hardware: Ethernetkabel (E), FireWire-Geräte (FW), Kopfhörer (K), Netzteil (NT), Maus (M), USB-Geräte (USB) und Zweites Display (ZD)

- Den Zustand bestimmter Hardware: Accelerometer (ACC), Batteriestatus (BAT), CPU-Auslastung (CPU), Eingabe Idle (EI), Lautstärkeinstellungen (LE) und Umgebungslicht (UL)
- Netzwerkkonnektivität: Aktiver Wi-Fi Access Point (AW), Wi-Fi-Access-Points (WI), Aktive Netzwerk-Interfaces (AN), Bluetooth-Geräte (BT) und IP Adresse (IP)
- Informationen zu Betriebssystem nahen Prozessen: Anwendung im Fokus (AIF), Laufende Anwendungen (LA), Fenster im Fokus (FIF), Gemountete Laufwerke (GL) und Screensaver (S)
- Informationen über einzelne Anwendungen: Aktive Chats (AC), Email (EM), Kalender (KA) und Skype (SK)
- Informationen zur physische Präsenz von Personen, durch das im Laptop eingebaute Mikrofon (Sprachaktivität (SA)) und die Kamera (Gesichtserkennung (GE))
- Informationen welche aus den ESM-Daten extrahiert werden können: Aufenthaltsort (AO), Stunde (ST), Tageszeit (TZ), Wochenende (WE) und Wochentag (WT)

Die verwendeten Sensoren können weiterhin hinsichtlich der zurück gelieferten Sensorevents in vier Klassen eingeteilt werden: *1x1 Sensoren* liefern einen Wert pro Sensorevent (z.B. liefert der *Anwendung im Fokus* Sensor nur den Prozessnamen der aktuellen Anwendung im Vordergrund); *1xn Sensoren* liefern mehrere, semantisch unterschiedliche Werte (z.B. liefert der *Accelerometer* Sensor drei einzelne Werte für die x, y und z Beschleunigung des im Rechner verbauten Beschleunigungsmesser); *nx1 Sensoren* liefern eine Liste von Werten des selben Typs (z.B. liefert der *Laufende Anwendungen*-Sensor eine Liste aller laufenden Anwendungsprozesse); und schließlich liefern *nxm Sensoren* eine Matrix von Werten zurück (z.B. liefert der *Wi-Fi-Access-Points*-Sensor eine Liste der aktuell verfügbaren Access Points mit deren SSID, BSSID sowie der korrespondierenden Signalstärke (RSSI) zurück).

Sowohl Sensorevents als auch ESM-Daten werden in XML-Format auf die Festplatte geschrieben, wobei Zeichenketten (z.B. der Name des Anwendung im Fenster) zur Wahrung der Privatsphäre in einen nahezu eindeutigen Hashwert umgewandelt werden, der für das maschinelle Lernverfahren ausreicht.

2.3 Merkmalsextraktion und -generierung

Die gesammelten Daten werden zunächst für das maschinelle Lernen vorbereitet. Dazu werden die Sensorevents für jeden einzelnen Probanden zusammengeführt, daraus Merkmale extrahiert und anschließend neue Merkmale generiert und die entstandenen Merkmalsvektoren, mittels der ESM-Daten, mit Labels versehen. Bei der Extraktion der Merkmale werden insbesondere die *1xn*, *nx1* und *nxm* Werte in ihrer Dimensionalität reduziert. Für den *Laufende Anwendungen*-Sensor (*1xn*) bedeutet dies beispielsweise, dass jede auftretende Anwendung in ein binäres Merkmal verwandelt wird. Danach wird für ein bestimmtes Sensorevent für alle in dessen Liste vorkommenden Anwendungen das entsprechende Merkmal auf *true* gesetzt.

Bei der Generierung von neuen Merkmalen werden die vorher extrahierten Merkmale über die Zeit integriert. Das heißt, ausgehend vom Erfassungszeitpunkt eines ESM-Datums werden jeweils die Sensorevents der davor liegenden x Minuten (5, 8 und 14) herangezogen und die Merkmale für diese Zeiträume analysiert. Binäre Merkmale werden dabei hinsichtlich der folgenden fünf Kategorien überprüft: keine der Werte (noneT), mind. ein Viertel der Werte (oqT), mind. die Hälfte der Werte (hlfT), mind. drei Viertel der Werte (tqT), und alle Werte (allT) der letzten x Minuten sind true. Daraus ergeben sich dann die einzelnen generierten Merkmale wie z.B. Application.Firefox.5m.noneT. Für numerische Merkmale werden jeweils Standardabweichung und Mittelwert für die letzten x Minuten berechnet. Der komplette Merkmalsvektor wird dann mit dem aus dem ESM-Datum gewonnenem Label in eine Datei geschrieben, die dem maschinellen Lernverfahren als Datensatz dient.

3 Durchführung und inhaltliche Auswertung

Die Daten wurden in einer vierwöchigen ESM Studie von vier Probanden (männlich, 25-33 Jahre) auf deren mobilen Rechner gesammelt, eingerahmt von einem kurzen Vorabfragebogen und strukturierten Interviews im Nachlauf. Die Stichprobengröße von vier Personen hat sich in vergleichbaren Arbeiten (Horvitz et al. 2004) als geeignet hinsichtlich der Balance aus Aussagekraft und Durchführbarkeit gezeigt. Drei der Probanden nutzten ihren Rechner während der gesamten Studiendauer, sowohl im Arbeitskontext als auch im privatem Umfeld; ein Proband nur im Arbeitskontext. Im Durchschnitt sammelten die Nutzer 338,2 (SA=171,2) ESM-Daten zusammen mit 270.000 (SA=133.612) Sensorevents.

Alle Probanden hatten zuvor regelmäßig IM genutzt (drei Probanden täglich, einer 3-4 Tage die Woche) und hatten 4-15 Jahre Erfahrungen mit IM-Diensten gesammelt. Dabei hatten sie 2-6 IM-Dienste—darunter immer Skype—mit im Durchschnitt 61 (Standardabweichung (SA)=23,4) Kontakten über alle IM-Dienste genutzt. Mit 11,2 (SA=2,5) davon hatten sie in den davor liegenden drei Monaten kommuniziert.

Die erste inhaltliche Auswertung zeigte, dass die Probanden spezifische Bedürfnisse nach selektiver Verfügbarkeit hatten. Während für einen Proband 100% seiner ESM-Daten selektive Verfügbarkeit zeigten, hatten für die restlichen drei nur 18,3% (SA=3,9%) der ESM-Daten unterschiedliche Verfügbarkeitslevels für die drei Verfügbarkeitskategorien. Analysiert man für diese drei Probanden die Daten nach Aufenthaltsort, sieht man, dass an einigen Orten das Bedürfnis nach selektiver Verfügbarkeit deutlich höher liegt als die durchschnittlichen 18,3%, während an anderen das Bedürfnis wiederum bei 0% liegt. Diese Zahlen wurden auch in den nachfolgenden Interviews bestätigt, in denen die Probanden ein ortsabhängiges Bedürfnis nach selektiver Verfügbarkeit zum Ausdruck brachten.

Letztlich sei noch auf die Differenz der Verfügbarkeitswechsel zwischen der durch ESM erfassten generellen Verfügbarkeit und der vom Skype Sensor erfassten Verfügbarkeit hingewiesen. Während die ESM-Daten in 31,6% (SD=11,4%) der Fälle einen Wechsel zeigten, waren es in Skype nur 6,9% (SD=3,8%).

Zusammenfassend lässt sich zu den Daten sagen, dass für unsere Probanden selektive Verfügbarkeit von Relevanz ist, und die manuelle Einstellung einen Mehraufwand darstellt.

4 Machbarkeitsanalyse der Vorhersagbarkeit

In diesem Abschnitt analysieren wir die Machbarkeit der automatischen Adaptierung ausdifferenzierter Onlinestatus für selektive Verfügbarkeit. Im Folgenden stellen wir die Auswahl der Algorithmen und die Ergebnisse der Evaluierung der Vorhersagbarkeit von Verfügbarkeit auf Basis der gesammelten und aufbereiteten Ergebnisse vor. Anschließend gehen wir darauf ein, welche Sensoren für die Vorhersagen ausschlaggebend waren.

4.1 Auswahl der Algorithmen für die Merkmalsselektion und Klassifikation

Das Verfahren zur Merkmalsselektion sowie der Algorithmus für die Konstruktion der Vorhersagemodelle hängen voneinander ab, da die verschiedenen Algorithmen unterschiedlich gute Vorhersagemodelle, abhängig von der jeweiligen Menge an Merkmalen, liefern. Daher wurden unterschiedliche Kriterien zur Merkmalsselektion verwendet (Information Gain, Gain Ratio, Chi Squared Metric, Consistency Subset, Correlation-based Feature Selection), um eine Reihe von Datensätzen mit einer jeweiligen Auswahl an Merkmalen zu erstellen. Auf Basis dieser wurde mit den in Frage kommenden Klassifikationsalgorithmen (Bayes Net, Sequential Minimal Optimization, C4.5 Decision Tree, Random Forest) Vorhersagemodelle erstellt und anschließend mittels der zehnfachen Kreuzvalidierung evaluiert. Als Basis dienten dabei die jeweiligen Implementationen der Algorithmen im WEKA Toolkit (Witten & Frank 2006). Die Ergebnisse zeigen dass das Merkmalsselektionsverfahren Correlation-based Feature Selection (CFS) in Kombination mit dem Algorithmus Sequential Minimal Optimization (SMO) in Form einer Support Vector Machine für unser Vorhaben am besten geeignet sind und folglich ausgewählt wurden

4.2 Evaluation der Vorhersagemodelle

Im ersten Schritt der Analyse wurde die Vorhersagbarkeit der generellen Verfügbarkeit untersucht. Hierzu wurden—wie in Absatz 2.3 beschrieben—entsprechende Datensätze für alle Studienteilnehmer erstellt. Mittels des CFS wird aus den Merkmalsvektoren dann die optimale Merkmalsmenge ausgewählt.

Wie sich gezeigt hat, sind die Datensätze stark unbalanciert. Das heißt, es existieren für jede mögliche Klasse (in diesem Falle Verfügbarkeitslevel) unterschiedlich viele Instanzen bzw. Beispiele, mittels welcher das Vorhersagemodell trainiert werden kann. Insbesondere Instanzen von Klassen, für welche nur wenige Trainingsinstanzen vorhanden waren, werden häufig falsch durch das resultierende Vorhersagemodell klassifiziert. Aus diesem Grund werden die Daten mit Hilfe des von WEKA bereitgestellten Resampling Filters ausbalanciert (Quinero-Candela et al. 2009).

Die so vorbereiteten Datensätze dienen als Eingabe für den ausgewählten SMO. Die gemittelten Ergebnisse für die Vorhersagbarkeit genereller Verfügbarkeit, ermittelt durch zehnmaliges Durchführen einer zehnfachen Kreuzvalidierung, sind in Tabelle 1 dargestellt.

Tabelle 1. Analyseergebnisse der Vorhersagbarkeit genereller Verfügbarkeit für Proband 1-4.

	Basis	Ergebnis	Differenz
P1	28.37	72.03	43.66
P2	23.43	93.55	70.12
P3	36.91	91.11	54.2
P4	28.25	85.3	57.05
Durchschnitt	29.24	85.49	56.25

Die durchschnittliche Basiswahrscheinlichkeit beträgt 29,24% (SA=5,60%). Dieser Wert gibt Aufschluss darüber, wie gut ein Klassifikator wäre, der stets die Majority-Klasse als gegeben annimmt. Das heißt in diesem konkreten Fall, dass etwa 29% der Fälle korrekt klassifiziert werden, sofern immer die am stärksten vertretene Klasse angenommen wird.

Die trainierten Vorhersagemodelle für generelle Verfügbarkeit leisten durchschnittlich eine korrekte Klassifikation in 85.49% der Fälle (SA=9.62%). Das bedeutet, der Unterschied zur Basis beträgt in den untersuchten Fällen im Durchschnitt 56.25% (SA=10,88%).

Tabelle 2. Analyseergebnisse der Vorhersagbarkeit der Verfügbarkeit hinsichtlich der drei Verfügbarkeitskategorien für Proband 1-4.

	Basis	Ergebnis	Differenz
P1-Private	28.37	75.38	47.01
P1-Work	21.53	75.97	54.44
P1-Public/Others	35.82	85.27	49.45
P2-Private	17.03	90.96	73.97
P2-Work	23.43	87.37	63.94
P2-Public/Others	17.03	94.15	77.12
P3-Private	36.91	90.95	54.04
P3-Work	36.91	95.37	58.46
P3-Public/Others	25.00	90.18	65.18
P4-Private	28.25	89.96	61.71
P4-Work	23.21	86.73	63.52
P4-Public/Others	28.25	84.95	56.7
Durchschnitt	26.811	87.27	60.46

Im zweiten Schritt der Analyse wurde die Vorhersagbarkeit von selektiver Verfügbarkeit spezifisch nach den drei Verfügbarkeitskategorien untersucht. Dazu wurde für jede Verfügbarkeitskategorie eines jeden Teilnehmers ein Trainingsdatensatz erstellt. Dies erlaubt die Untersuchung der Vorhersagbarkeit einer jeden Verfügbarkeitskategorie. Auch hier wurde wieder mittels CFS die am besten geeignete Menge an Merkmalen ausgewählt. Und auch diese Datensätze werden mittels des Resampling Filters von WEKA ausbalanciert, sodass die

verschiedenen Klassen (z.B., *Offline*, *Do Not Disturb*, *Away*) mit nahezu gleich vielen Instanzen trainiert werden können.

Die Ergebnisse der Evaluierung (zehnmaliges Durchführen einer zehnfachen Kreuzvalidierung) sind in Tabelle 2 dargestellt, wobei die Ergebnisse für die drei Verfügbarkeitskategorien einer jeden Versuchsperson gruppiert sind. Die Basiswahrscheinlichkeit für die korrekte Erkennung der jeweiligen Verfügbarkeit beträgt im Durchschnitt 26,81% (SA=6,99%). Die Erkennungsleistung der trainierten Vorhersagemodelle beträgt im Durchschnitt 87,27% (SA=6,28%). Die Erkennungsleistung für die verschiedenen Verfügbarkeitskategorien ist unterschiedlich und weist keine Muster auf, sodass die Schlussfolgerung nahe liegt, dass die Vorhersagbarkeit der Verfügbarkeit von Verfügbarkeitskategorien von den individuellen Gegebenheiten der jeweiligen Versuchsperson abhängt. Diese Ergebnisse deuten darauf hin, dass Vorhersagbarkeit auch für eine Vielzahl von BenutzerInnen im ähnlichen Maß möglich ist.

4.3 Einfluss der Sensoren auf die Vorhersagemodelle

Für die Integration von Vorhersagemodellen in Anwendungen ist es von Interesse, welche Sensoren die wertvollsten Merkmale beigetragen haben. Sieben Sensoren (BAT, EI, EM, FIF, LA, UL, WI) waren für alle vier Teilnehmer von Bedeutung. Das heißt, dass diese Sensoren für Merkmale verantwortlich sind, die zur Vorhersage der Verfügbarkeit beitragen. Acht Sensoren (ACC, AW, AIF, IP, LE, SA, ST, USB) hatten für jeweils drei Teilnehmer eine hohe Aussagekraft und fünf (FW, GL, K, M, S) hatten für keinen Teilnehmer eine Auswirkung auf das Modell. Das heißt, es gibt Sensoren, die Parameter erfassen, die für das Erstellen der Vorhersagemodelle einzelner BenutzerInnen von großer Bedeutung sind, für andere allerdings nicht. Für Sensoren, deren Daten bei der Merkmalselektion nicht berücksichtigt wurden, ist zu beachten, dass deren Informationsgehalt eventuell von anderen Sensoren bereits abgedeckt wurde.

5 Verwandte Arbeiten

Im Folgenden gehen wir auf Arbeiten ein, die sich ebenfalls mit der Vorhersagbarkeit von Benutzerselbsteinschätzungen auf Basis von Sensordaten beschäftigen. Da nach unserem Wissen keine Arbeiten existieren, welche sich mit der Voraussage von selektiver Verfügbarkeit beschäftigen, betrachten wir zwei prominente Arbeiten, welche sich mit der Vorhersage von genereller Unterbrechbarkeit beschäftigen.

Auch Fogarty et al. (2004) nutzen ESM zur Erfassung der Unterbrechbarkeit auf einer 5-stufigen Likert-Skala. Die Probanden wurden dabei über den Rechnerlautsprecher zu einer mündlichen Aussage (1-5) bezüglich ihrer Unterbrechbarkeit aufgefordert. Mittels ca. 7 Sensoren werden Daten der Desktoprechner (Maus, Keyboard, aktive und inaktive Anwendungen) sowie der Büros (Bewegung, Telefon, Sprache, Tür) aufgezeichnet. Dabei sammelten sie 975 Benutzerselbsteinschätzungen von 10 Probanden (2 Manager, 5 Forscher, 3 Praktikanten). Die mittels des naive Bayes Klassifikators erstellten Modelle, welche die Daten

aller Subjekte mit einbeziehen, erreichten eine Vorhersagegüte von 79,5%. Modelle, für einen Probandentyp (z.B. Praktikant) zwischen 80,1% bis 87,7%. Für diese Klassifikation wurde allerdings das Model auf zwei Klassen reduziert (1-4 und 5). Der Unterschied zu unserer Arbeit besteht weiterhin in zwei Bereichen. Erstens, der Einsatz von mobilen Sensoren führt bei uns zu personalisierten Modellen, die individuell pro NutzerIn trainiert werden, dadurch allerdings eine höhere Vorhersagegüte erreichen. Zweitens wurde in der beschriebenen Arbeit keine selektive Verfügbarkeit bzw. Unterbrechbarkeit untersucht.

Auch das *BusyBody* System (Horvitz et al. 2004) basiert auf der ESM, um die Kosten von Unterbrechungen bei BenutzerInnen zu messen und deren Vorhersagbarkeit zu analysieren. Die vier Teilnehmer konnten dabei über die Dauer mehrerer Wochen in einem präsentierten Dialogfenster zwischen ‚Beschäftigt‘ und ‚Nicht beschäftigt‘ wählen. Die Daten wurden mit einem ähnlichen Set an Sensoren wie bei Fogarty et al. kombiniert und mittels Bayes’scher Netze die Voraussagbarkeit für jeden Probanden separat berechnet. Horvitz et al. erreichen dabei eine durchschnittliche Vorhersagegüte von 78,25%, allerdings auch nur für eine binäre Klassifikation zwischen ‚Beschäftigt‘ und ‚Nicht Beschäftigt‘—gegenüber der komplexeren Vorhersage von sechs Verfügbarkeitslevels in unserer Arbeit. Weiterhin analysieren Horvitz et al. ebenfalls nicht die Vorhersagbarkeit von selektiver Verfügbarkeit bzw. Unterbrechbarkeit.

6 Zusammenfassung und Ausblick

In dieser Arbeit präsentierten wir Untersuchungsergebnisse zur Vorhersagbarkeit von selektiver Verfügbarkeit mittels maschinellen Lernens. Die in unserer Studie erreichte hohe Vorhersagegüte für die vier Probanden zeigt die generelle Machbarkeit solcher Vorhersagemodelle. Der Einsatz solcher Vorhersagemodelle für IM-Systeme kann zukünftig den BenutzerInnen ein größere Ausdifferenziertheit hinsichtlich ihrer Erreichbarkeit bei gleichzeitig sinkendem Benutzeraufwand ermöglichen.

Inwiefern automatisch generierte Vorhersagen von BenutzerInnen angenommen werden, sollte in künftigen Benutzerstudien gezeigt werden. Des Weiteren wäre es auch interessant zu beleuchten, wie sich solche Systeme durch kontinuierliches Lernen an plötzliche oder schleichende Veränderungen (Concept Drift und Concept Shift) anpassen können. Um tatsächlich einsetzbare Systeme zu realisieren, ist schließlich noch eine Herausforderung, Trainingsdaten implizit aus der Nutzung des Systems zu sammeln und dabei die BenutzerInnen nicht zusätzlich zu unterbrechen und ihre Privatsphäre zu wahren.

Literaturverzeichnis

Csikszentmihalyi, M. & Larson, R. Validity and Reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease* 175, 9 (Sept. 1987), pp. 526-536.

- Davis, S. & Gutwin, C. Using Relationship to Control Disclosure in Awareness Servers. In Proceedings of Graphics Interface 2005 - GI 2005 (9.-11. Mai, Victoria, British Columbia). A K Peters Ltd., 2005. pp. 145-152.
- Fetter, M., Gross, T. & Zeller, B. Disclosure Templates: Vorlagen fuer die selektive Freigabe personelicher Informationen. In Mensch & Computer - 8. Fachuebergreifende Konferenz fuer interaktive und kooperative Medien - M&C 2008 (7.-10. Sept., Lübeck, Deutschland). Oldenbourg, 2008. pp. 57-66.
- Fogarty, J., Hudson, S.E., Akteson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. & Yang, J. Predicting Human Interruptibility with Sensors. ACM Transactions on Computer-Human Interaction (TOCHI) 12, 1 (Mär. 2005). pp. 119 - 146.
- Fogarty, J., Hudson, S.E. & Lai, J. Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. In Proceedings of the Conference on Human Factors in Computing Systems - CHI 2004 (24.-29. Apr., Wien, Österreich). ACM Press, 2004. pp. 207-214.
- Horvitz, E., Koch, P. & Apacible, J. BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work - CSCW 2004 (6.-10. Nov., Chicago, Illinois). ACM Press, 2004. pp. 507-510.
- Kapoor, A. & Horvitz, E. Principles of Lifelong Learning for Predictive User Modeling. In Proceedings of the 11th International Conference on User Modeling - UM 2007 (25.-29. Jul., Korfu, Griechenland). Springer, 2007. pp. 37-46.
- Mileswski, A.E. & Smith, T.M. Providing Presence Cues to Telephone Users. In Proceedings of the 2000 ACM Conference on Computer-Supported Cooperative Work - CSCW 2000 (2.-6. Dez., Philadelphia, PE). ACM Press, 2000. pp. 89-96.
- Patil, S. & Kobsa, A. Instant Messaging and Privacy. In HCI 2004: 18th British HCI Group Annual Conference (6.-10. Sept., Leeds, UK). Springer, 2004. pp. 85-88.
- Quinero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N.D., Hrsg. Dataset Shift in Machine Learning. MIT Press, Cambridge, MA, 2009.
- Voida, A., Newstetter, W.C. & Mynatt, E.D. When Conventions Collide: The Tensions of Instant Messaging Attributed. In Proceedings of the 2002 Conference on Human Factors in Computing Systems - CHI 2002 (20.-25. Apr., Minneapolis, MN). ACM Press, 2002. pp. 187 - 194.
- Witten, I.H. & Frank, E. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2006.

Danksagung

Wir danken den Mitgliedern des Cooperative Media Lab sowie den Studienteilnehmern. Das Projekt wird gefördert durch das BMVBS und den Projektträger Jülich unter dem Förderkennzeichen FKZ 03WWTH018.

Kontaktinformationen

Prof. Dr. Tom Gross, tom.gross(at)medien.uni-weimar.de, T. 03643/58-3733