Combining data mining and text mining for detection of early stage dementia: the SAMS framework

Christopher Bull, Dommy Asfiandy, Ann Gledson, Joseph Mellor, Samuel Couth, Gemma Stringer, **Paul Rayson**, Alistair Sutcliffe, John Keane, Xiaojun Zeng, Alistair Burns, Iracemi Leroi, Clive Ballard and Pete Sawyer
Lancaster University, University of Manchester, King's College London

Overview & key points

- SAMS project and framework

- Combination of data and text collection and analysis

- Non-invasive method

- Features linked to deficits in cognitive domains characteristic of dementia

- Previous papers on requirements gathering and ethical concerns, focus here is on implementation of text and data collection components

Dementia UK context

- 1 in 6 people at the age of 80

- Predicted to increase from 850K to 2M by 2051

- Diagnosis of dementia or MCI usually performed using paper-based cognitive texts such as MoCA

- Problems of referral until disease is well advanced

Research context

- Growing body of research and interest in health-related research
  - RaPID-2016 workshop @ LREC
  - Three "Computational Linguistics and Clinical Psychology" workshops held annually at ACL/NAACL since 2014
  - Six "International Workshops on Health Text Mining and Information Analysis" held at various locations since 2008
  - NIPS 2015 Workshop on Machine Learning in Healthcare

- Previous data not easily available
  - Nun-study
  - Iris Murdoch, Agatha Christie, PD James
  - DementiaBank

- Detect early signs of dementia, even before people are aware of problems

- Three-year project: monitoring computer-use activity to effectively detect subtle signs of cognitive impairment

- Promoting self-awareness of change in cognitive function

- School of Computing and Communications, Lancaster University, UK

- School of Computer Science, University of Manchester, UK

- Institute of Brain, Behaviour and Mental Health, University of Manchester, UK

- Wolfson Centre for Age-Related Diseases, King's College London, UK

SAMS — Software Architecture for Mental Health Self-Management

- Project aims
  - Non intrusive capture of computer use
    - Text (e.g. email), mouse movements, keyboard typing.
  - Mine the data for trends and patterns
    - mapping to clinical indicators e.g. working memory, motor control.
  - Infer longitudinal changes in cognitive health
   - possible early onset of dementia

- Internet activity provides an opportunity to monitor for changes over time

- SAMS will validate optimal thresholds by examining changes in performance

- Feedback mechanisms to enhance users' cognitive self-awareness

- Generic user-centered feedback architecture which could be applied to a wide range of problems

SAMS | Software Architecture for Mental Health Self-Management

Desktop/Application
Monitor Processes
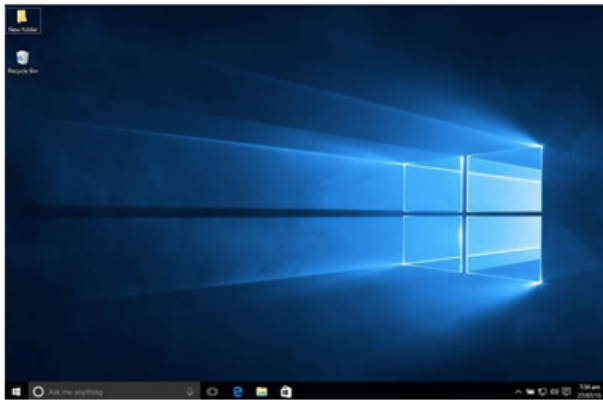
Encrypt
Logs

Browser Extensions

Manager Process

Secure SAMS
Server

- Desktop logger component
  - Level 1: Keyboard and mouse

    Application and Global Mouse and Keyboard Hooks .Net library in C#

  - Level 2: Operating system (e.g. desktop activities)
    - FileSystemWatcher class
    - Clipboard (.Net)
    - Microsoft UI Automation events
  - Level 3: Applications
    - Office Primary Interop Assemblies
    - Internet Explorer Object
  - Combination of above data to create higher level events, e.g.
    - Phases of mouse movements and drag phases
    - Mapping of desktop icons/windows

- Web browser extensions … some major issues
  - Applications are largely black boxes
  - Ensuring privacy/security (Ethics)
  - Dynamic webpages (use JS MutationObserver)
  - Other engineering or usability considerations

https://mail.google.com/mail/u/3/#inbox?compose=154c48b169cb4190

Google

Gmail ▾

More ▾

1–2 of 2

COMPOSE

Inbox
Starred
Important
Sent Mail
Drafts (1)
More labels ▾

Primary

Social

Promotions

Google   New sign-

GitHub   [GitHub]

0 GB (0%) of 15 GB used
Manage

Christopher ▾

No recent chats
Start a new one

Regarding today's meeting
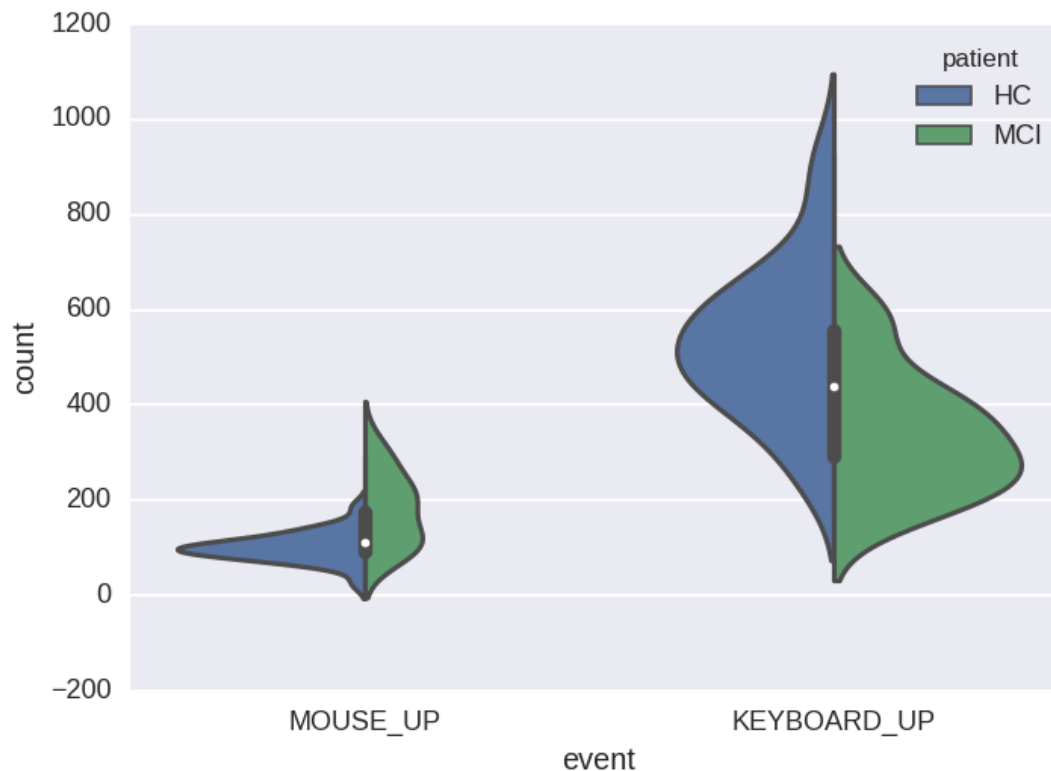
example@example.com

Regarding today's meeting

Dear Dr John Smith,

Thank you for talking with me today after the meeting.

I wish to take you up on your suggestion and look into working together on Project X.

Kind regards,
Jon Doe

Send

- Differences between groups with low-level mouse and keyboard events. (Significant with respect to Mann-Whitney U test)

Conclusions

- Novel system architecture that combines data (keyboard, mouse and OS/application) and text (email, diary events)

- Analyse change over time

- Longitudinal study will end in August

- Currently implementing text mining features from related work in Wmatrix analysis pipeline (lexical, grammatical and semantic)

- Framework will be available from UCREL GitHub

- For further information: http://ucrel.lancaster.ac.uk/sams/