Activity-logging for self-coaching of knowledge workers

Saskia Koldijk
Radboud University Nijmegen
The Netherlands
saskia.koldijk@student.ru.nl

Mark van Staalduinen
TNO
Delft, The Netherlands
mark.vanstaalduinen@tno.nl

Stephan Raaijmakers
TNO
Delft, The Netherlands
stephan.raaijmakers@tno.nl

Wessel Kraaij
Institute for Computing and
Information Sciences
Radboud University Nijmegen
w.kraaij@cs.ru.nl

ABSTRACT

With an increased societal focus on a sustainable economy and a healthy population, well-being of knowledge workers has become an important topic. This paper investigates techniques to support a knowledge worker to manage his well-being. A possible solution is to monitor the workers' behaviour and use this information for giving feedback as soon as his well-being is threatened. Knowledge workers use a broad range of communication means to achieve their goals, like a computer and mobile phone. Our research aims at using features like mouse clicks, active applications or key presses, because these are rather simple features to obtain instead of more invasive tools like a heart-rate monitor. This paper presents the first results of our research. First, logging of low-level features is developed. Based on these features the behaviour of different users is investigated. At first sight, this behaviour seems to be rather chaotic, but by taking into account different tasks, more structure is observed within the data. This paper shows that different behaviour is observed for different users and different tasks, while the same characteristics are observed when a user is performing the same task. This suggests that also anomalous behaviour might be recognized, which is an important result for developing self-coaching tools.

1. INTRODUCTION

In the modern knowledge economy, the demands for productivity of knowledge workers are steadily increasing. At the same time, information sources and communication means are more fragmented than ever. Real-time communication means, such as e-mail, (micro)blogging and other social media have generated an overflow of information, lacking a structure that is adapted to the user's tasks. Networked information systems and portable devices make it possible to work anywhere, posing challenges to context aware net centric organisation of documents, task lists etc. Finally, since the work-force in Western countries is ageing it is increasingly important to develop supportive techniques that help people having a reduced work capacity due to a medical condition to maintain a healthy work-style. The project User Centric Reasoning for Well-working (UCR4W)[1] is investigating the key determinants for well-being at work. One of the guiding hypotheses of the project is that logging the activities of knowledge workers can be the basis for an effective computer based coach. The objective of the project is to develop user-centric sensing and reasoning techniques that help to improve well-being at home and at work. Technology should help people improve their sense of being and feeling in control, with a positive impact on work efficiency and effectiveness, work pleasure, mental and physical health status. An example of empowerment is to have relevant information from personal data collections available 'just-in-time'. We think that understanding the activities and tasks of individuals is a key condition to achieve this.

In this paper we describe a study that is carried out as a precursor to the UCR4W project and present initial results of an experiment. The underlying idea of the study is that knowledge workers could possibly be helped to adapt their work style by providing them neutral feedback about their work style and activities. Section 2 discusses the background and assumptions of the study. Experimental results are presented in Section 3.

2. FEEDBACK FOR SELF-COACHING

The information overload and context switches of knowledge workers can be a threat for productivity and well-being at work. Let us consider the following scenario:

Scenario

Imagine a typical working day as a knowledge worker. You have different projects running and today your plan is to work on project A and B. For project A you do some internet search and start typing a document. While you are busy a colleague asks your help for project C. You inter-

[1]UCR4W is a 8 MEuro 4 year project co-funded through the Dutch national research programme COMMIT. Partners are: TNO, Novay, University of Twente, Radboud University Nijmegen, Philips research, Ericsson, Roessing R&D, Noldus Information Technology, Netherlands Centre for Social Innovation, Dutch Ministry of the Interior.

rupt your work to help her searching for some information. When a mail about project B arrives you decide to switch to this project as it is quite urgent to finish the required document. Suddenly you notice it is 5 o' clock and you did not finish your work on project A as planned. You might start wondering: How did I spend my time today?

The relation between work-style and well-being at work

Knowledge workers typically have various tasks and deadlines and they have to produce results. The possibility to easily switch tasks makes working very fragmented. The course of action is not always self planned but also determined by external causes, like phone calls, mails, information requests, other persons or appointments [3]. Knowledge workers typically have to self-manage their work and make a good planning in order to be able to accomplish all their tasks. All in all this way of working easily causes a feeling of stress and it is quite difficult to keep a good overview what it is one has done over the course of a day, weeks or even months. A study has shown that knowledge workers often spend effort in tracking their own tasks (13%). Automating this process would be of great benefit for the working process. A system that could monitor and provide overviews of performed activities could support the worker with his self-management, adapt his work-style [1] and in this way diminish cognitive load and stress. More awareness of ones own working process might also have beneficial effects on the on-task behavior and adherence to scheduled activities [9].

Feedback based on action recognition

As a first step to test the hypothesis that tracking activity and work-style can improve well-being at work a simple feedback tool is under development which can automatically infer and log the tasks a user is performing. The log information could be presented in the form of a daily or weekly overview, showing the amount of time spent on tasks and the number of interruptions or task switches. Such a tool requires the following steps: i) design of an activity ontology, ii) automatic logging of low level computer interaction data iii) developing an inference module that maps low level activity to the activity ontology level iv) developing an effective presentation mode for feedback purposes. In this paper we report work on the first three steps, the main contributions of our study are related to i) and iii). The first step necessary in this research is the creation of a taxonomy of tasks people could be performing. Several taxonomies of tasks have already been proposed in the literature. The taxonomies about internet use by Morrison, Pirolli and Card [7] indicate that a distinction between actions on three different levels might be appropriate: The method the user adopts, the purpose of his actions and the specific content. The next step of this project is task recognition. A model will be made for the inference process from simple logging data to higher level tasks. We intend to compare several types of features, both static and temporal in combination with various classifier schemes. In section 3 we will report initial work, since the experiments are still ongoing. A final step is connecting the tasks recognition module to a graphical user interface. It is important to make the interaction with the system as pleasant as possible. Myers et al. [8] state that the system should be directable, personalizable, teachable and transparent. So in order to work well the system should optimally cooperate with – and adapt to – to the user. The tool should provide a means for the user to give feedback without irritating him and it should keep learning.

Related work

There has already been done much theoretical and applied research in the field of action understanding (e.g. [2, 4, 11]). Research on pattern recognition in sensor data, multimodal fusion and models for human goal directed behaviour are relevant for our work. Some research has specifically focused on recognizing patterns of user activities on pc's [6, 10]. These studies all focus on the detection of a specific kind of information to trigger certain actions. Our research differs as we want to log all kinds of activity in order to make a human understandable overview and categorization of tasks. Our intent is to give the user a better overview and more awareness about his working process and in this way help him improving his performance.

3. PILOT STUDY: WORK-LOGGING

3.1 Task analysis

Following a user centred approach, a questionnaire was used to investigate the typical way of working of knowledge workers and their demands on software supporting them in their daily practices. From the responses by 47 knowledge workers at TNO we concluded that for a tool being usable for support, the captured activity data should be aggregated to a higher level in order to provide the user with valuable information. The recognition of the task a user is performing is a useful first step towards providing the user understandable feedback and insights about his working process. On the basis of the questionnaire a set of tasks that knowledge workers perform was identified. The answers to the questions 'What tasks do you perform and how do you use your computer for this?' and 'Describe a typical working day' were manually grouped into sets of similar answers to derive a set of typical task types. The appropriateness of the identified set of task types was confirmed by several knowledge workers. From all task types, the tasks performed at the computer were finally selected for automatic task recognition (see Table 1 for the tasks labels used).

3.2 Data collection

After identification of the software demands by the users, our next step consisted of investigating whether the computer could possibly fulfil these demands. In an experimental phase the computer activities of three knowledge workers were logged using uLog[2]. An additional tool was created that reminded every 10 minutes to annotate his activity by selecting one of the labels from the task list (and indicating his level of wellworking). About two weeks of data collection resulted in a labelled raw data set. The labelled raw dataset was processed to extract several features, for example how often the user clicked or which application was mainly in focus within a five minute time frame (cf. Table 3 for a full list of extracted features; cf. Table 1 for the amount of data points per label). In total 20, 180, 66 labelled segments were recorded for the three users respectively.

[2]http://www.noldus.com

Table 1: Dataset - amount of data per label

Task label	# data	as percentage	F-value
read mail	11	4%	0.583
write mail	12	5%	0.348
organize archive data	5	2%	0
plan	14	5%	0
make presentation	3	1%	1
create visualisation	4	2%	0.857
program	63	24%	0.977
write report paper	82	31%	0.8
search information	17	6%	0.654
read article text	17	6%	0.746
make overview	31	12%	0.621
analyse data	7	3%	0
TOTAL	266	100%	0.656

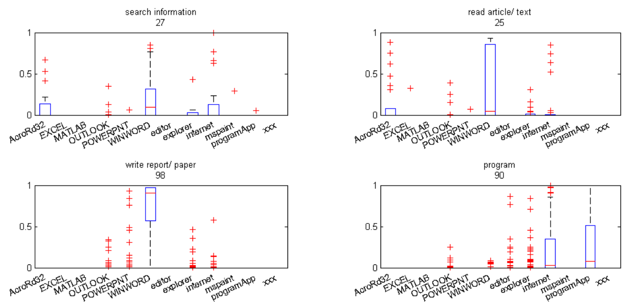

Figure 1: Distribution of application usage (as percentage of time that the application was in focus) per task

3.3 Analysis of the labelled data

First analysis showed that distinguishable patterns of computer activity arose per assigned task label. The most indicative feature seems to be the application that was mainly in focus, which is logical as specific tasks require specific applications, as for example 'programming' is done in a programming application. Also the distribution of all applications in the time frame seems to be useful. For both the tasks 'write report' and 'search information' Word has main focus, but someone 'searching for information' additionally uses an Internet browser and AcrobatReader (see Figure 1).

Besides the used applications the keyboard and mouse activity can be used to further distinguish tasks. Figure 2 shows the distribution of clicks and typed characters for the different task labels. Some features alone already have discriminative power (see Figure 3 for an indication of information gain ratio per feature), for example the amount of typed characters is about 0 for searching information, about 50 for mail writing and about 200 for report writing. Combining more features increases the discriminative power, for example tasks not discriminable by number of typed characters (for example writing mail and making an overview, both about 50 typed characters) could be recognized on basis of the number of clicks (about 40 vs. about 80).

A final useful feature that could indicate the task a user is performing is the amount of switching between different applications. Figure 3 plots the typical distribution for various users to show that there are clear individual differences.

3.4 Experiment: Automatic activity labelling

Some initial results about automatic activity labelling are available (see Table 2). We used Weka (see Hall et al. [5]) to train some classifiers and tested their performance by means of 10 fold cross validation. Labelling each activity simply as the majority class 'write report/ paper' with Weka's ZeroR classifier yielded a baseline accuracy of 30.83% (F=0.145). Using Weka's Naive Bayes classifier with just the feature mainApp to classify tasks resulted in an accuracy of 59.77% (F=0.468), so we can conclude that the application that was mainly in focus is a very strong feature. Adding the other features with mouse and keyboard information and indications about active applications and application switches (discretized with Weka's preprocessing option) improved the classification accuracy to 70.30% (macro-averaged F=0.656; F values per task can be found in Table 1). Leaving out all features that address the use of specific applications, classification accuracy drops to 52%, with an average F=0.45, which stresses that application-dependent information is important as well for task identification.

Table 2: Classification - initial results

Classifier	Accuracy	Averaged F
Baseline (classify as main class)	30.83%	0.145
Naive Bayes (use only mainApp)	59.77%	0.468
Naive Bayes (use all features)	70.30%	0.656

4. CONCLUSION AND FUTURE WORK

The pilot experiment suggests that rather different behaviour is observed for different users, while they are performing the same task. This implies that self-coaching should be personalized tooling. Furthermore, it is concluded that task recognition is essential to monitor the well being of a knowledge worker, since low level information is not informative.

The dataset needs to be extended to several dozens of users, in order to draw more substantial conclusions. The data collection will probably be extended with a component that captures some semantic content that helps to model the interaction of well-being with an activity related to a particular project.

The next steps within the UCR4W project aim at the recognition of anomalous behaviour for each task that might signal a decreasing well-being of a worker. Furthermore, the project will evaluate the self-coaching tools together with end-users in order to improve its acceptance. Finally, proper privacy protection mechanisms and procedures will be an integral part of the project, because these tools are based on personal data.

5. REFERENCES

[1] C. Argyris and D. Schön. *Organizational learning: A theory of action perspective.* Addison Wesley, 1978.
[2] C. L. Baker, R. Saxe, and J. B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329 – 349, 2009. Reinforcement learning and higher cognition.
[3] M. Czerwinski, E. Horvitz, and S. Wilhite. A diary study of task switching and interruptions. In *CHI '04: Proceedings*

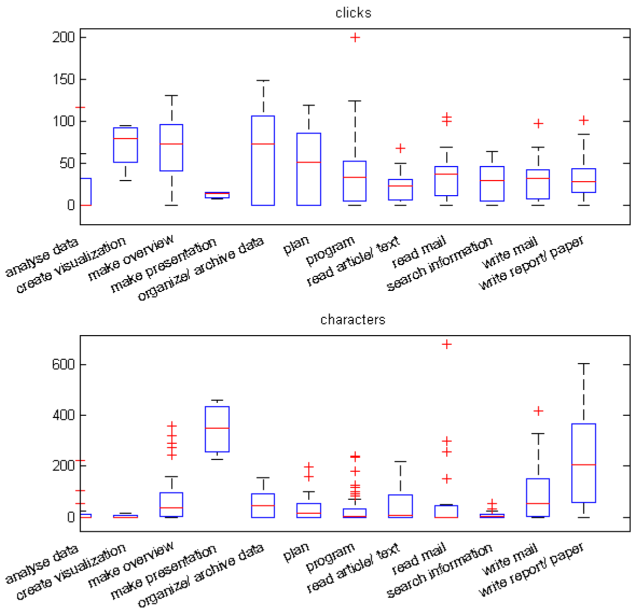

Figure 2: Distribution of #clicks and #characters per task

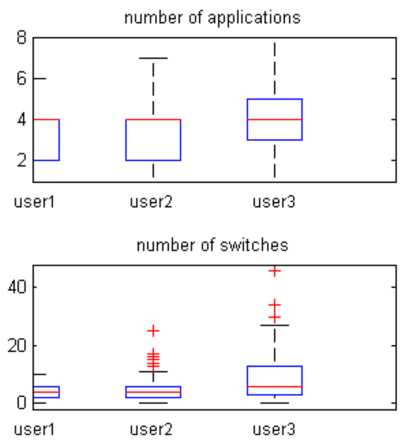

Figure 3: Application usage and switching behavior per user

of the SIGCHI conference on Human factors in computing systems, pages 175–182, New York, NY, USA, 2004. ACM.

[4] T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. volume 1, pages 838 – 845 vol. 1, jun. 2005.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009.

[6] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and D. Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265. Morgan Kaufmann Publishers, July 1998.

[7] J. B. Morrison, P. Pirolli, and S. K. Card. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 163–164, New York, NY, USA, 2001. ACM.

[8] K. Myers, P. Berry, J. Blythe, K. Conleyn, M. Gervasio, D. McGuinness, D. Morley, A. Pfeffer, M. Pollack, and M. Tambe. An intelligent personal assistant for task and time management. *AI Magazine*, 28(2):47–61, 2007.

[9] G. S. Richman, M. R. Riordan, M. L. Reiss, D. A. Pyles, and J. S. Bailey. The effects of self-monitoring and supervisor feedback on staff performance in a residential setting. *J Appl Behav Anal.*, 21(4):401Ű409, 1988.

[10] J. Shen, L. Li, T. G. Dietterich, and J. L. Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces*, pages 86–92, New York, NY, USA, 2006. ACM.

[11] K. Tahboub. Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. *Journal of Intelligent & Robotic Systems*, 45:31–52, 2006. 10.1007/s10846-005-9018-0.

Table 3: Features extracted within 5 minute timeframes, sorted by information gain ratio (GR)

Feature name	Description	GR
mspaint	% time that mspaint had focus	0.847
user	the user who logged and labelled the data	0.765
programApp	% time that a programming application had focus (eclipse, cmd...)	0.677
OUTLOOK	% time that OUTLOOK had focus	0.626
WINWORD	% time that WINWORD had focus within the timeframe	0.597
mainApp	application that was most of the time in focus	0.522
AcroRd32	% time that AcroRd32 had focus	0.472
spaces	# spaces typed	0.39
characters	# characters typed	0.364
backspaces	# backspaces (inc. 'delete'-key)	0.361
specialKeys	# special keys typed	0.344
clicks	# clicks within the timeframe	0.29
switches	# switches between applications	0.26
internet	% time that an internet application had focus (iexplorer, firefox...) within the timeframe	0.251
daytime	time of the day (as hour, i.e. 9-18)	0
scrolls	# scrolls	0
nrApps	# different applications used within the timeframe	0
time	% time this mainApp was in focus	0
editor	% time that an editing application had focus (notepad++, wordpad...)	0
POWERPNT	% time that POWERPNT had focus	0
explorer	% time that the explorer had focus	0
EXCEL	% time that EXCEL had focus	0
MATLAB	% time that MATLAB had focus	0
label	task label for the activity given by the user	