Supervision of learning methods in user data interpretation

Maya Sappelli
TNO & Radboud University Nijmegen
m.sappelli@cs.ru.nl
Supervisors: Suzan Verberne and Wessel Kraaij

1. MOTIVATION

Knowledge workers need support in their daily activities to handle the information stream they encounter continuously (e.g. e-mail, search results etc.). One method for this is to place the information in context, i.e. to what activity is the information related? For this purpose, contexts need to be recognized. Additionally, this should happen in an unobtrusive way such that the worker is not disrupted and can remain focused at his work-related activities. We formulated the following research questions:

- How can we create a user model of activities and interests by observing the user's interaction with his personal information devices and how do we measure its quality?
- How can we use an automated assistant to provide information support to knowledge workers to increase their feeling of being in control and increase productivity without being obtrusive?
  - When is it appropriate to ask a user for supervision on the model learn process?
  - How can we balance obtrusiveness and functionality of the intelligent assistant? (i.e. when should we push relevant information and when should we wait for the user's initiative?)

The PhD project is titled "User centered content filtering".

2. METHODOLOGY TO BE USED

We propose to recognize context by monitoring the interactions between user and computer. This includes low-level activities such as key presses and mouse movements. The recognized context (e.g. the project for which a document is accessed) can be used to provide the user with relevant information, as well as insight into his work activities.

Research by Koldijk et al.[1] showed that the task activity of the user (e.g. writing an e-mail, reading an article) can be reliably determined from these computer interaction patterns. However, it is dependent on the user which classifier is the most effective predictor. The results showed that only 2.5 hours of training data was enough to achieve 80% recognition accuracy.

3. INITIAL RESULTS

Research such as the work by Koldijk et al.[1] requires users to provide training examples for the classifiers by labeling examples or time slots with the associated task activity or task content. One method to reduce the required input from users that these classifiers need, is by using existing information on possible categories (i.e. categorized documents) to categorize unstructured data (i.e. e-mails). Our experiments [2] showed that the documents do not have the same characteristics as the e-mails which makes them not suitable as training examples.

4. PLANNED NEXT STEPS

The initial results indicated that at least for categorizing e-mails it is likely that the user will need to provide training examples himself. Therefore, we focus on reducing the number of training examples needed. We plan to explore the possibility of combining supervised and unsupervised learning methods. Additionally, we are searching for the best way to ask supervision from the user at appropriate times using appropriate methods.

For the remainder of the thesis, we have two overall goals. The first is to increase the user's feeling of control by giving him a better overview of his activities. The second is to increase the user's productivity by helping him organize the information flow on his system and safeguarding his workflow by delaying irrelevant notifications.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] S. Koldijk, M. van Staalduinen, M. Neerincx, and W. Kraaij. Real-time task recognition based on knowledge workers' computer activities. In *Proceedings of ECCE 2012 (Edinburgh, Scotland)*, 2012.

[2] M. Sappelli, S. Verberne, and W. Kraaij. Using file system content to organize e-mail. In *Proceedings of the fourth symposium on Information interaction in context*, 2012.