Eye Tracking-Based Desktop Activity Recognition with Conventional Machine Learning

by Ole Poeth

supervised by Lingyu Du and Guohao Lan

1. Background

- **Context sensing** through **eye movement** has shown a convincing correlation[1], making gaze signals a potential candidate to use for **activity recognition**.
- An ideal set of **eye movement features** for classification has not been found yet, while different sets of features have shown great potential for accurate classification.
- This research will look into **eye movement features** of gaze signals and the impact of **subject** bias on the classification.

2. Research Questions

1. How to design and implement different feature extraction methods for eye movement signals?
2. What are the best features that need to be extracted and used for training conventional machine learning algorithms?
3. What's the impact of different subjects and sensing hardware on the recognition performance?
4. Compare deep and conventional machine learning algorithms on accuracy and robustness against heterogeneity among subjects.

3. Method

In order to get to the classification of gaze data different steps are performed to get the desired results. The methodology consists of the following steps:

- Data from **8** different subjects performing **5** different tasks on a computer while wearing an **eye tracker**.
- The data consists of **9000** points each resembling the relative position of the gazes. This is pre-processed with a **median filter** and then normalized.
- A **fixation filter**[2] is used to extract **fixations** and **saccades** from the data.
- From these saccades and fixations specific features are extracted and the **most important** features are used for analysis.

Category	*Sub-Category*	*Features*
Fixation	Duration	fix-dur-mean, fix-dur-var, fix-dur-std
	Rate	fix-rate
	Slope	fix-slope
	Dispersion Area	fix-disp-area
	Radius	fix-radius
Saccade	Length	sac-len-mean, sac-len-var, sac-len-std
	Direction	sac-dir-nne, sac-dir-ene, sac-dir-ese, sac-dir-sse, sac-dir-ssw, sac-dir-wsw, sac-dir-wnw, sac-dir-nnw

Table 1. This table shows all the features that are extracted from the fixations and saccades.

- During classification the data is divided into a **75/25** training-test. For the testing of subject bias, the split is performed in two different ways: on **activities** and on **subjects**.

4. Results

Tables 1, 2 and 3 show the classification accuracy results of three different tests: **best feature set** using feature importance, **subject bias** by splitting data on activities and subjects, comparison between **deep learning** and **conventional machine learning** algorithms.

Feature Sets / *Classifiers*	*All*	*Fixation*	*Saccade*	*Best*
SVM	0.95	0.59	0.88	0.95
k-NN	0.84	0.49	0.81	0.86
Random Forest	0.92	0.63	0.86	0.92

Table 1. This table shows the results of the different features sets on three different conventional machine learning algorithms. Fixation features do not perform well on their own, saccade features do show great importance, but all features minus fixation radius perform the best.

Classifiers	*Data Split On Activities*	*Data Split On Subjects*
SVM	0.95	0.60
k-NN	0.84	0.54
Random Forest	0.92	0.58

Table 2. This table shows the results of the classification of three different classifiers when the data is split on activities and when on subjects. There is a clear drop in performance when the data is split on subjects.

	SVM	*LSTM*	*CNN*
Activities	0.95	0.95	0.97
Subjects	0.55	0.32	0.38

Table 3. This table shows the results of the best performing conventional machine learning classifier(SVM) and two deep learning classifiers(LSTM and CNN) on the same data set split on activities and on subjects. Deep learning classifiers show minor to no performance increase on activity split but a greater drop on subject split.

5. Conclusions

The conclusions for this research are based on the research questions asked at the beginning and are summarized per question:

1. Fixation and saccade are the best gaze features to extract from the data, the fixation filter gives control on extraction for better results.
2. Saccade features show good results on their own, while fixation features are dependent. A combination shows the best results up to 95% accuracy.
3. Heterogeneity of data plays a significant role in classification performance where the difference in accuracy between classifying on activities and on subjects is up to 35%.
4. Deep learning classifiers do not show huge improvement in comparison to conventional machine learning algorithms but do take significantly longer to train, making them not a obviously better than conventional classifiers.

References

[1] Lan, Guohao, et al, "GazeGraph: graph-based few-shot cognitive context sensing from human visual behavior", Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020.
[2] Olsson, Pontus, "Real-time and offline filters for eye tracking." (2007)