Universität Rostock

Traditio et Innovatio

Recognition of Emotions based on Sensor Data

KSWS / Project by **Ole Adelmann, Mhd Esmail Kanaan, Dennis Kelm, Jonny Schlutter** and **Gia Huy Hans Tran**

Introduction

Emotion recognition (ER) through the use of AI is starting to see use in various fields. ER is most commonly derived from audio and/or video data, but is also possible using motion-sensor data. However, as success with sensor data on its own is limited, there are approaches trying to improve performance through the use of external knowledge. Identifying predispositions in subjects can increase human activity recognition (HAR) performance substantially, as shown by Popko et al. [1].

In this project we applied similar methods to identify *emotional* predispositions (EPDs). We implemented a data-driven pipeline originally used for HAR by Popko et al. [1] and adapted it for the task of ER using techniques from Sultana et al. [3] concerning data preparation and construction of a robust emotion model.

We will give a brief overview on how emotional predispositions (EPDs) are identified using clustering and how they can be used to improve our results when classifying emotions. During evaluation we compare different classifiers, amounts of clusters (i.e., EPDs), and additionally general versus personal models.

Extracting EPDs

We used clustering to group intervals, where a person has had similar mood over all days. By dividing up each day into small time segments and calculating the distribution of emotions over each segment (Histograms, see *Fig 4.1*), we can map them out as shown in *Fig. 4.2*. Now, the task is to divide these up into groups, so that points that lie close to each other are in the same group. This is called clustering.

In *Fig. 4.2*, you can see that the points are partitioned by color. Each color represents a cluster identified by the *K-Means* clustering algorithm, and each of these clusters is one of our EPDs.

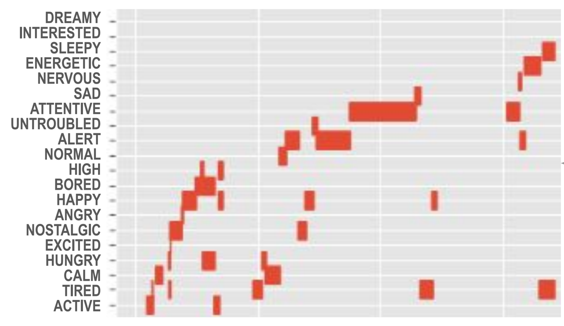


Fig. 1.1 - Annotations of some of the 49 moods

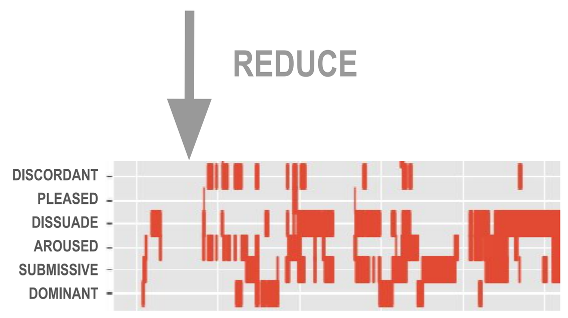


Fig. 1.2 - Reduction to 6 emotional states

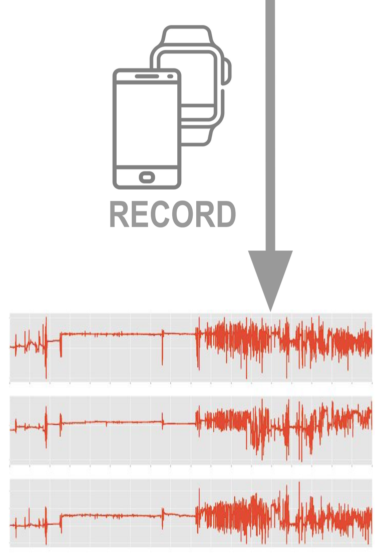


Fig. 2.1 - Motion Sensor Data [2]

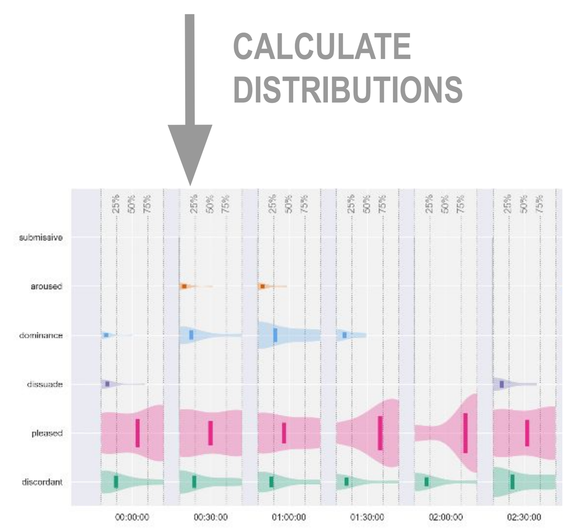


Fig 4.1 - Histograms of Person 21, 0:00-2:30

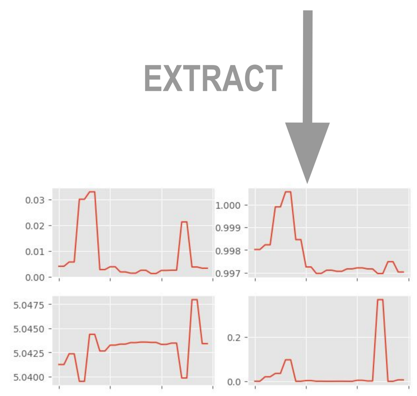


Fig. 2.2 - Features

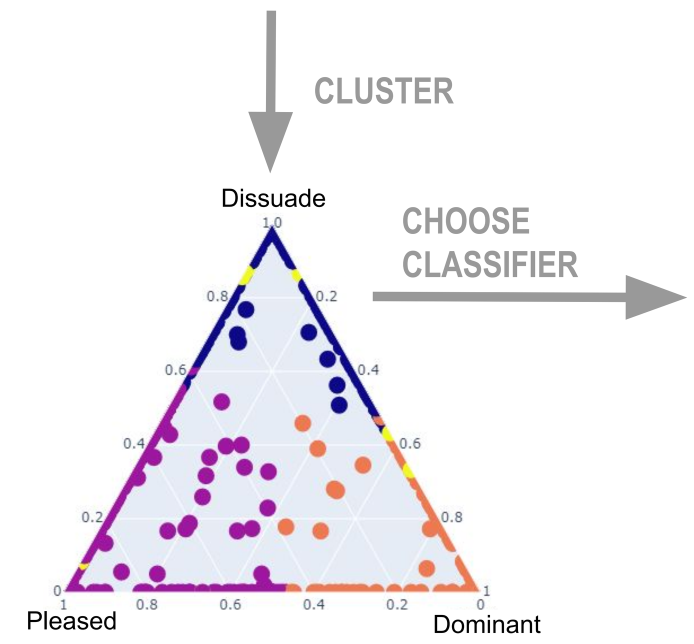


Fig. 4.2 - Histograms mapped out and clustered

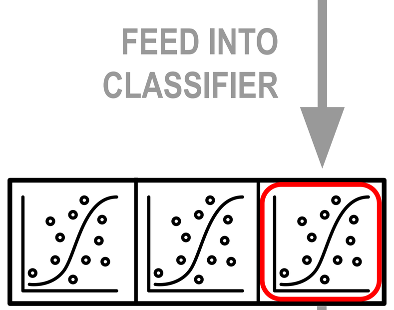
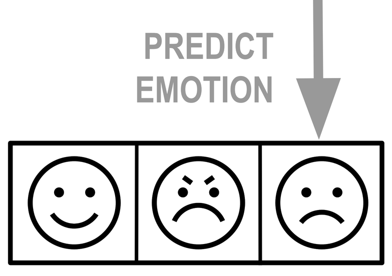


Fig. 5 - Model containing a classifier for each EPD



Data and Preparation

We trained our models using the *ExtraSensory dataset* [2]. It consists of sensor data gathered from smartphones and smartwatches of 60 participants. The data was recorded in intervals of one minute and was labeled with activities and mood.

Our selected features are 138 statistical and spectral measures (*Fig. 2.2*), extracted by Vaizman et al. [2], from their accelerometer, gyroscope, and magnetometer data (*Fig. 2.1*). (e.g., their means, stds, magnitude spectrums)

There were a total of 49 different discrete mood labels (*Fig 1.1*). Using the PAD method used by Sultana et al. [3], we reduced this down to 6 emotional states (*Fig. 1.2*).

Out of the 60 subjects, we selected 18 who had more than 1000 samples and noted their mood on more than 10% of their samples (*Fig. 3*).

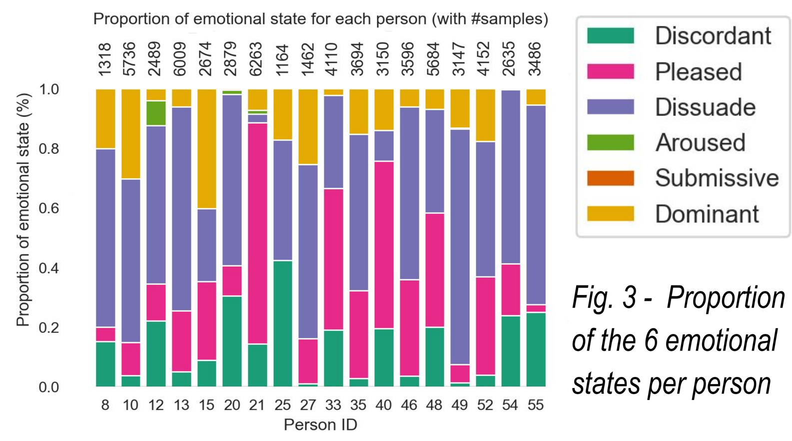


Fig. 3 - Proportion of the 6 emotional states per person

Classification of Sensor Data

Now that we have our EPDs, we train a classifier for each of them. Classification is the task of learning to map an input to a desired output. In this case, we want to map motion sensor data onto one of 6 emotions. By training a classifier on a certain EPD rather than on the entire dataset, we can make use of a person's tendencies and improve the accuracy of our predictions. Our model m as a whole (*Fig. 5*) takes sensor data s and an EPD d, and chooses a classifier c_d, which maps s to an emotion e.

$$m(s,d) = c_d(s) = e$$

Experimental Evaluation

Experimental Design

For each of our 3 different classification methods (Majority, Support Vector Machine (SVM), Random Forest), we compared: a single generalized model (GM) vs. models for each person (PM), with the amount of clusters ranging from 1 (i.e. no clustering) to 10. We evaluated on four metrics (see *Fig. 6*), using stratified 5-fold cross validation.

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	proportion of correct predictions
Precision	$\frac{TP}{TP+FP}$	fraction of relevant instances among the retrieved instances
Recall	$\frac{TP}{TP+FN}$	fraction of relevant instances that were retrieved
F1 Score	$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$	harmonic mean of the precision and recall

Fig. 6 - evaluation metrics

Results

Personalizing the models always led to higher average Accuracy, F_1 and Precision, regardless of the classifier. Moreover, increasing the number of clusters has a greater impact on PMs than on GMs.

Compared to our other methods, Majority classification (our baseline) performed poorly when using PMs, but was surprisingly accurate on GMs. The Support Vector Machine (SVM) is the least accurate for GMs but outperforms Majority classification for PMs. (*Fig. 7.2*)

For GMs, the Random Forest classifier performs similarly to SVM, but for PMs, it performs visibly better than Majority and even SVM, regardless of the number of clusters. The difference in accuracy between Random Forest and the second-best classifier can be as high as 10%. We see similar results with the F_1 and Precision scores. (*Fig. 7.1*)

The Recall score is relatively constant across most classifiers but drops with rising cluster amount for personalized Majority and SVM.

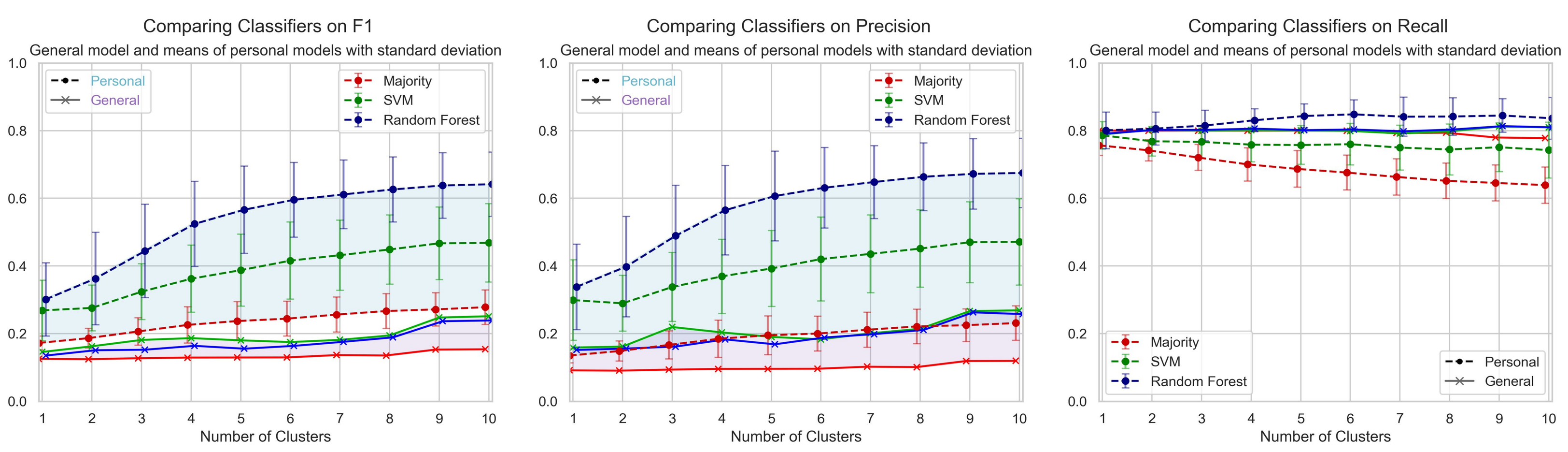


Fig. 7.1 - F_1 score, Precision and Recall of the classifiers compared

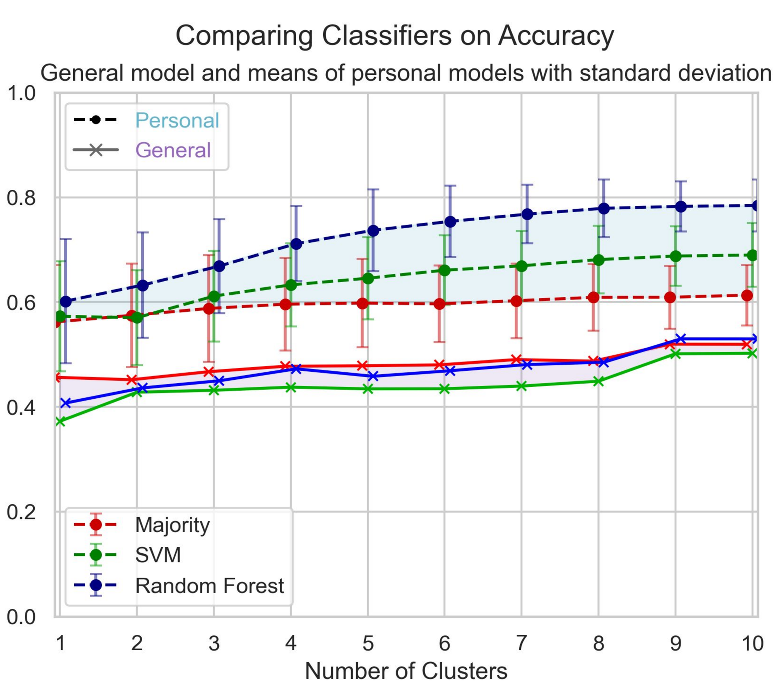


Fig. 7.2 - Accuracy of the classifiers compared

Conclusions

- ➔ EPDs identified through clustering can be utilized to improve Emotion Recognition (ER) performance.
- ➔ Personalized ER models exhibit higher average accuracy than generalized models.
- ➔ There is an upward trend in accuracy and F_1 as the number of clusters (EPDs) increases, with the improvement being much more prominent in personalized ER.
- ➔ The choice of the classifier is more impactful for personal ER than for the general ER.

References

[1] M. Popko, S. Bader, S. Lüdtke, and T. Kirste, "Discovering behavioural predispositions in data to improve human activity recognition," in *Proceedings of the 7th International Workshop on Sensor-based Activity Recognition and Artificial Intelligence*, 2023.

[2] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervasive Comput.*, vol. 16, no. 4, pp. 62–74, 2017. Image Source: The official website: http://extrasensory.ucsd.edu/

[3] M. Sultana, M. Al-Jefri, and J. Lee, "Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: Exploratory study," *JMIR MHealth UHealth*, vol. 8, no. 9, p. e17818, 2020.

Icons from Flaticon.com

Supervisor: Maximilian Popko
Faculty of Computer Science and Electrical Engineering
University of Rostock, Albert-Einstein-Str. 22
18059 Rostock, Germany