Thesis BA/MA: Exploratory Cluster Analysis of Smartphone Users based on their Mobile Sensing Data

Mobile Sensing apps that are installed on users’ smartphones can access many kinds of user behavioral and contextual data. Many recent studies use that data to understand users’ behavior. That ranges from smartphone related behavior, i.e. how app usage varies with context (Hossein Falaki, 2010) (Trinh Minh Tri Do, 2011) to real-world behavior like day-night behavior patterns (Ramona Schödel F. P., 2020) and relations of smartphone use to personality (Clemens Stachl, 2020).

However users are mostly treated as a homogenous group. To design successful interventions and adaptive systems it instead is necessary to regard the subtle differences between sub-populations (Karen Church, 2015). E.g. Jones et al. (Maalej, 2015) show that there are at least 3 different kinds of smartphone users regarding app revisitation patterns.

Recent research found a huge amount of user types, regarding only app usage data (Sha Zhao, 2016): Zhao et al. clustered users a.o. into night communicators, screen checkers, car lovers, ...

However app usage is just a small subset of the data modern Mobile Sensing applications are able to collect. The aim of this thesis is to find clusters of users based on a more general dataset, that includes more than just app usage. We therefore provide access to a longitudinal mobile sensing dataset (Ramona Schödel M. O., 2020) that encompasses smartphone behavioral (e.g. smartphone usage, mobility, music consumption, ...) and contextual (e.g. location, device sensor data, ambient noise and brightness, ...) log data (N=800, duration 3 to 6 months).

The thesis could comprise of the following steps:

- Literature research on existing work about clusters of smartphone users
- Choosing and designing higher-level features from the sensing data, based on literature research or other suitable methods
- Implementation of feature extraction in the statistics language R
- Research for appropriate statistics methods to cluster the data
- Implementation of that clustering and evaluation of the results
- Writing and presenting your thesis

Requirements:

- Independent working style and the ability to come up with own solution ideas to your tasks
- Passion for data wrangling and analytics, and first experiences in that topic
- Knowledge of the statistics language R and workflow with RStudio with be ideal, but not mandatory
- Willingness to work on LMU infrastructure in the lab at Frauenlobstraße 7a for some tasks for data security reasons (as far as the Covid situation allows)

As the scope and required skills for this thesis is relatively advanced, I would rather recommend it as master thesis than bachelor thesis. However if you feel confident and interested in that topic, it is also possible as Bachelors thesis.

If you are interested please contact florian.bemmann@ifi.lmu.de, including your transcript of records and a brief statement about your motivation and why you feel that this topic is suitable for you.
Literaturverzeichnis


