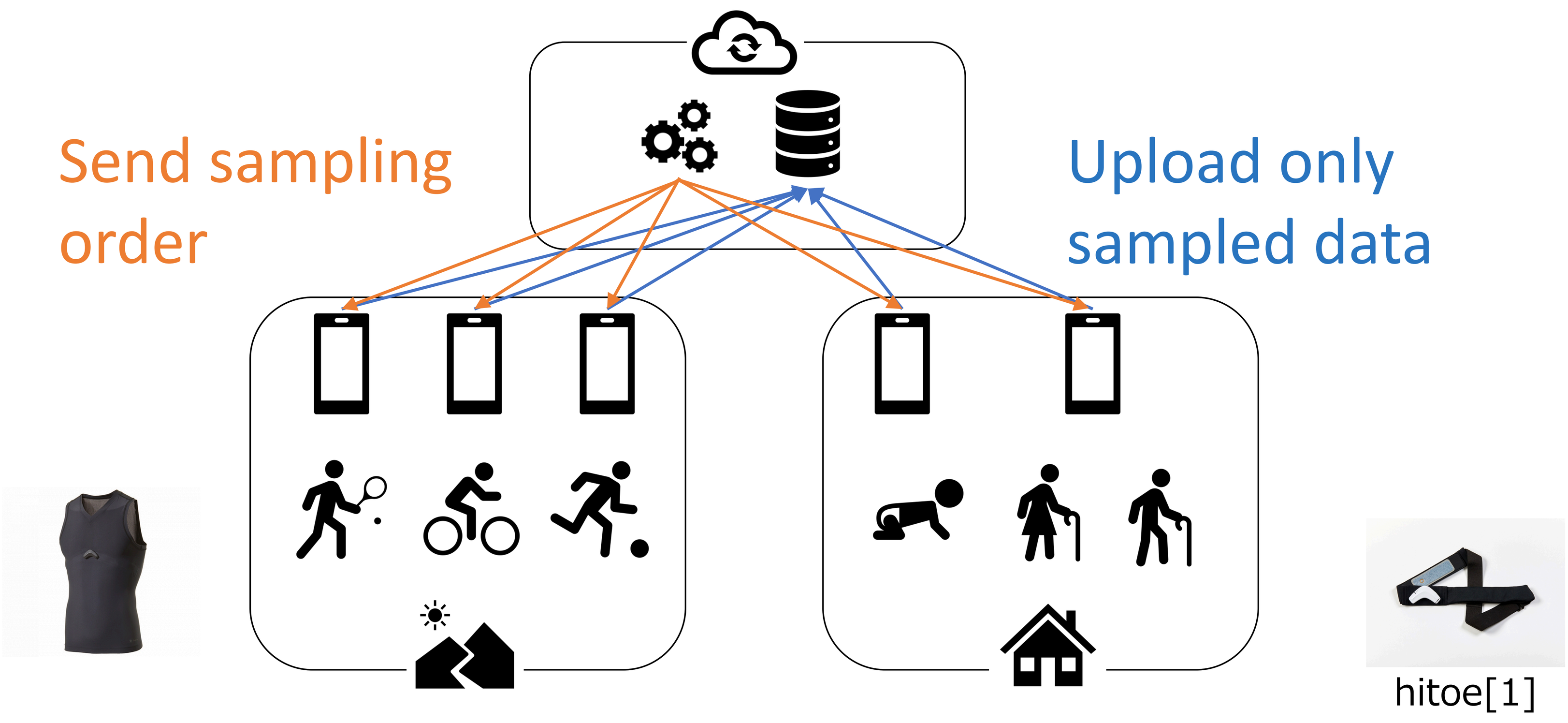
Adaptive Biosignal Data Gathering for Distributed and Continual Remote Monitoring

Osamu Saisho, Keiichiro Kashiwagi, Yui Saito, Tomoyuki Fujino (NTT)

Introduction

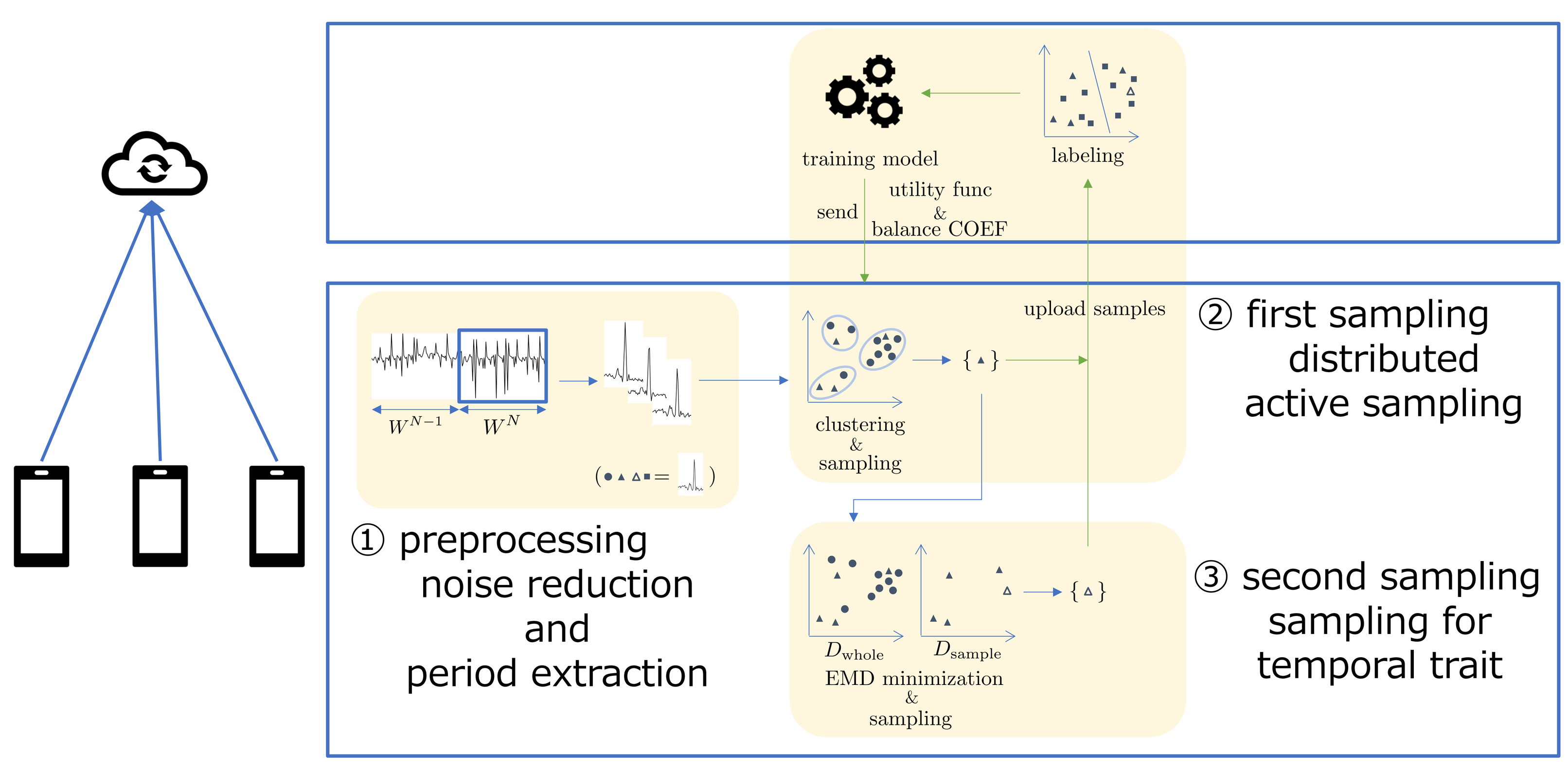
- Demand for remote biosignal sensing application is growing.
- Ideal apps requires
  - a high-performance machine learner
  - a dataset that shows the temporal trait of each user
- In a continual sensing environment, both need to be built simultaneously.
- However, uploading all data of all users is not desirable in terms of communication bandwidth and cloud capacity.
- We propose an distributed active sampling method for the remote sensing environment



Methods

- The requirements for distributed sampling in real systems are very strict.
  - Learning model in a cloud cannot see any data that was not uploaded.
  - Edge devices cannot see any other device's data.
  - Edge devices have only limited resources.
  - Data distribution changes over time.

- Three components of our proposed method
  - Preprocessing
    - General filtering and period extraction
  - First sampling (active sampling)
    - Acquisition function is derived from uncertainty sampling[2].
    - Training status in a cloud are propagated to edges through K-shape clustering[3].
  - Second sampling (user's temporal trait)
    - Additional data are sampled to minimize EMD[4] between sampled data and whole.

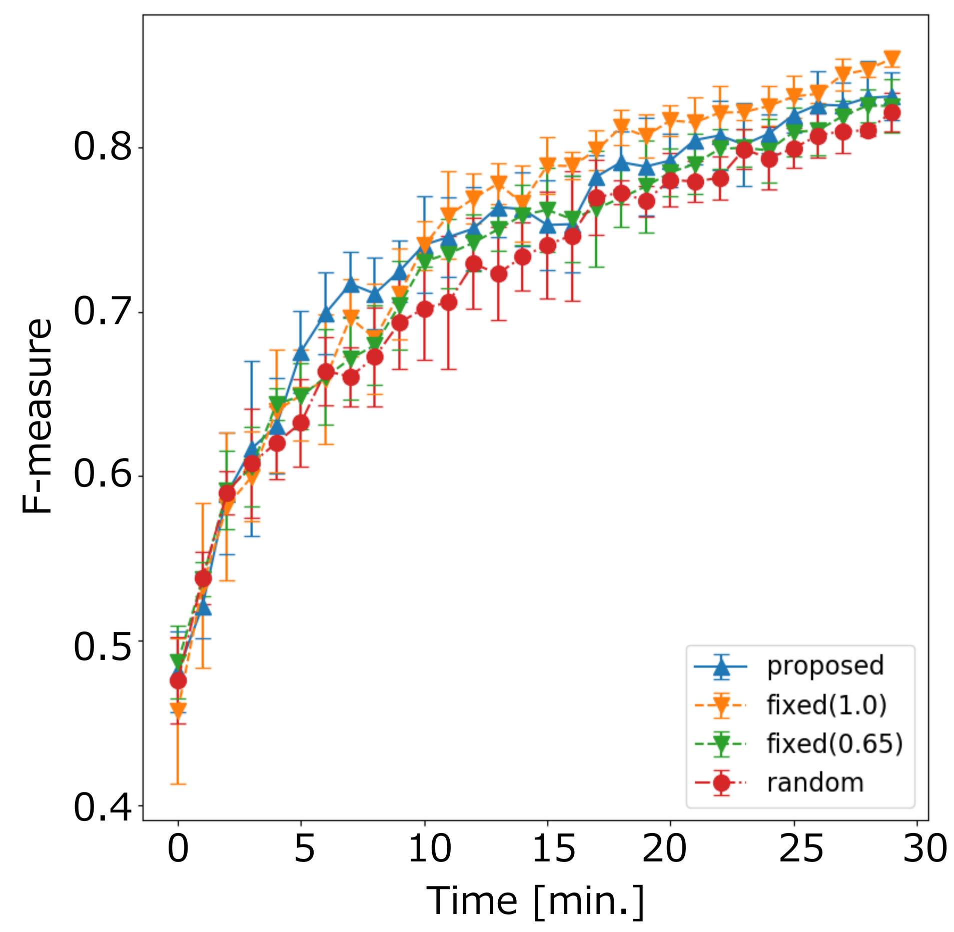


Experiment

- Experimental settings

dataset	MIT-BIH[5]
data description	five-class arrhythmia classification (N: 83.1%, V: 6.3%, S: 2.5%, F: 0.7%, Q: 7.4%)
data volume	about 30 min. ECG data from 48 individuals
classification model	1d-CNN[6]
sampling rate	six beats / min.
comparative methods	• proposed method • only active sampling (fixed(1.0)) • fixed(0.65) proposed method • random sampling

- Experimental results
  - Active sampling methods achieved higher performance than random sampling.
  - Proposed method achieved higher performance than fixed(0.65).



  - Proposed method achieved smaller trait deviation than active sampling.

Euclidian distances between histograms of the classes of sampled data and whole (*10^{-2})

proposed	fixed(1.0)	fixed(0.65)	random
6.52	7.73	5.17	2.35

Conclusion

- We proposed a distributed active sampling method for continual biosignal monitoring.
- The experimental results verify the effectiveness for model training, graspability for temporal user traits, and adaptability.
- Our future work is evaluating the method in a real system and adding other factors to the consideration.

References

[1] https://www.hitoe.toray/en/
[2] B. Settles, Active Learning Literature Survey, 2009
[3] J. Paparrizos and L. Gravano, k-shape: Efficient and accurate clustering of time series, SIGMOD2015
[4] N. Bonneel et al., Displacement interpolation using Lagrangian mass transport, TOG2011
[5] G. Moody and R. Mark, The impact of the MIT-BIH Arrhythmia Database, Eng. Med. Biol. Mag. 2001
[6] A. Hannun et al., Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 2019