Handling Data Annotation Uncertainty in Human Activity Recognition

Hyeokhyen Kwon, Gregory Abowd, Thomas Plotz

Motivation

- Data Collection and Annotation are important!
- Most HAR models used a discrete data labeling scheme
- But even the most reliable annotator can get it wrong
    - Exact activity transition points are hard to determine
    - Activity can overlap each other
- **Label jitter:** uncertainty region at the start/end of an activity (activity boundary)

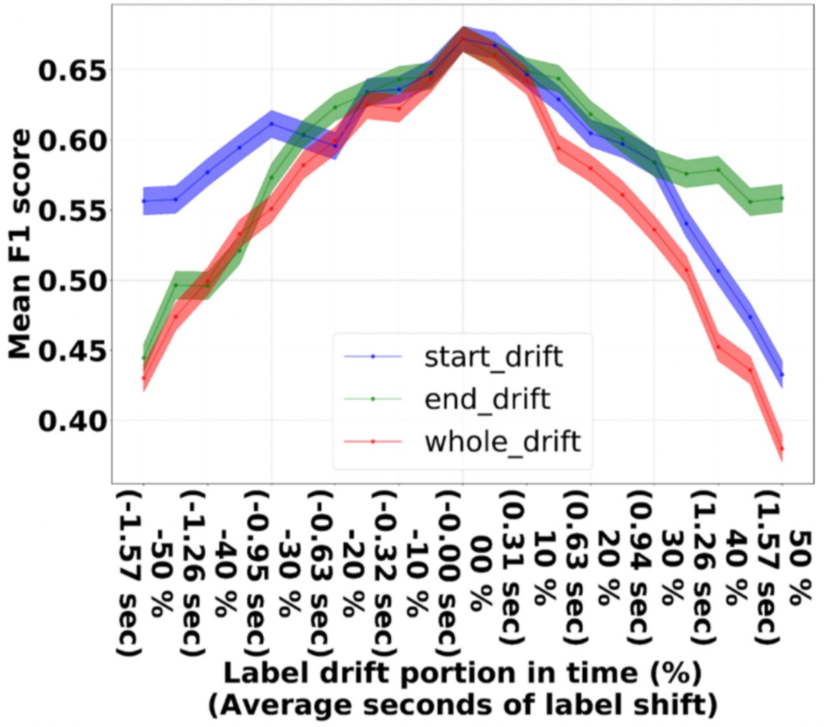
-> Label jitters can prevent the HAR model from achieving ideal performance

Past attempts

- **Discrete label correction**
  - Ignorance is bliss!
  - Majority voting among multiple annotators
  - Iteratively shift label to refine annotation (EM method, Fukunaga Class-Separability)
- **Soft label fusion**
  - Use multiple annotators -> subjectivity and uncertainty
  - Labels are represented as multinomial distribution of activities
    - Uniform contribution
    - Assign weights based on reliability
- **Label augmentation**: Incorporate noise into the dataset to regularize models

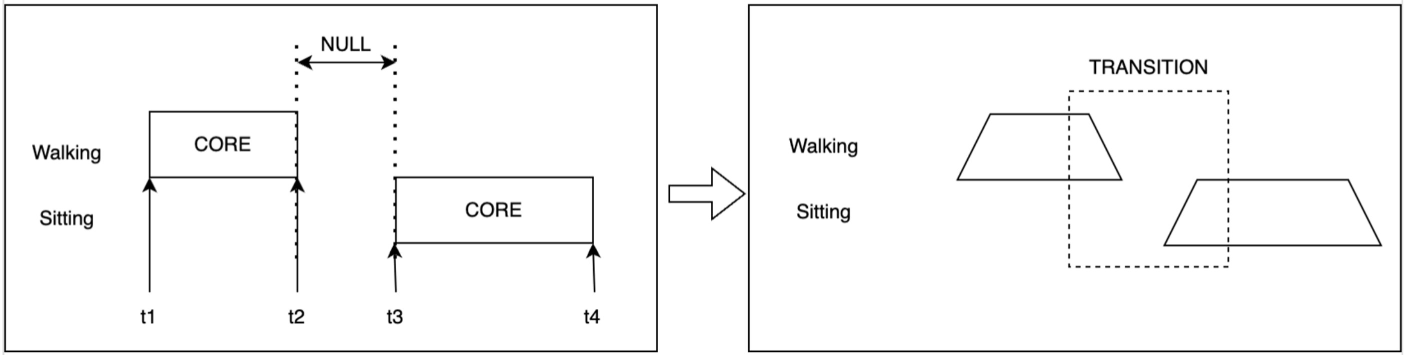
Effects of Label Misalignment on HAR models

- Experiment:
  - Opportunity dataset, ConvLSTM model
  - Shift the labels at the start/end/both of the boundary
  - Amount of shifting is proportional to the length of the activity (max 50%)
- Effects on the F1 score shown in the plot on the right



Adaptive Smooth Activity Transition (SAT)

- **CORE interval:** middle section of an activity, essence of the motions
- **NULL interval:** where nothing “interesting” happen
- **Event boundaries:** Start/End-points of an activity
- Perspectives:
    - Label jitter is unlikely at CORE interval
    - Activity Transition (TR) uncertainty inversely proportional to activity length
    - Overlapping (OV) uncertainty is inversely proportional to NULL interval length



Adaptive Smooth Activity Transition (SAT)

- Derive the smoothing kernel based on the observations:

$$P(t_b) = \gamma\mathcal{N}\left(t_b, \sigma_{t_b}^{TR}\right) + (1\ - \gamma)\mathcal{N}\left(t_b, \sigma_{t_b}^{OV}\right)$$

$$\sigma_{t_b}^{TR} = \frac{\alpha}{L_t} \text{ and } \sigma_{\mathrm{t_b}}^{OV} = \frac{\beta}{\Delta_t}$$

Constraint:

$0 \leq \gamma \leq 1$

$\sigma_{t_b}^{TR} \leq U^{TR} = L_t * p^{TR}$

$\sigma_{t_b}^{OV} \leq U^{OV} = (\Delta 1..t, p^{OV})$

$0 \leq p^{TR}, p^{OV} \leq 0.5$

Variables:

L is the length of the activity

Δ is the length of the NULL interval

α, β is the control parameters, with $[\alpha, \beta] = percentile([L_{1..t}, \Delta_{1..t}], 50\%)$

Smooth Activity Transition

Finally, we use P(t) as the kernel to smooth activity boundary at timestep t.

How can we use SAT with classifiers?

- Classifiers that work with soft labels: fuzzy SVM, softmax classifier, ...
- Discrete classifiers: augment the dataset by sampling from the uncertainty distribution

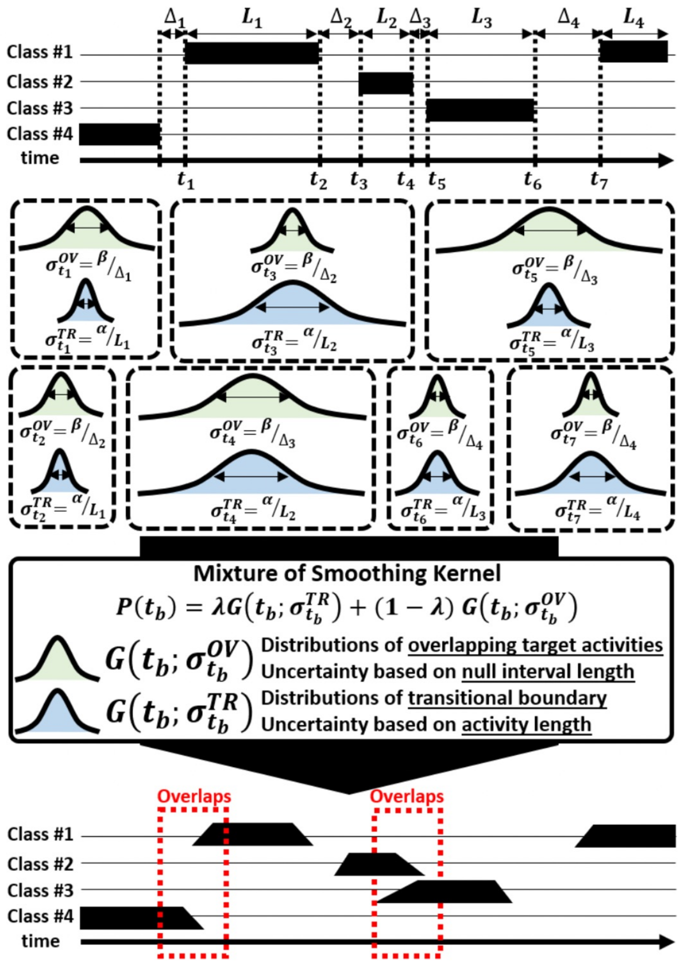
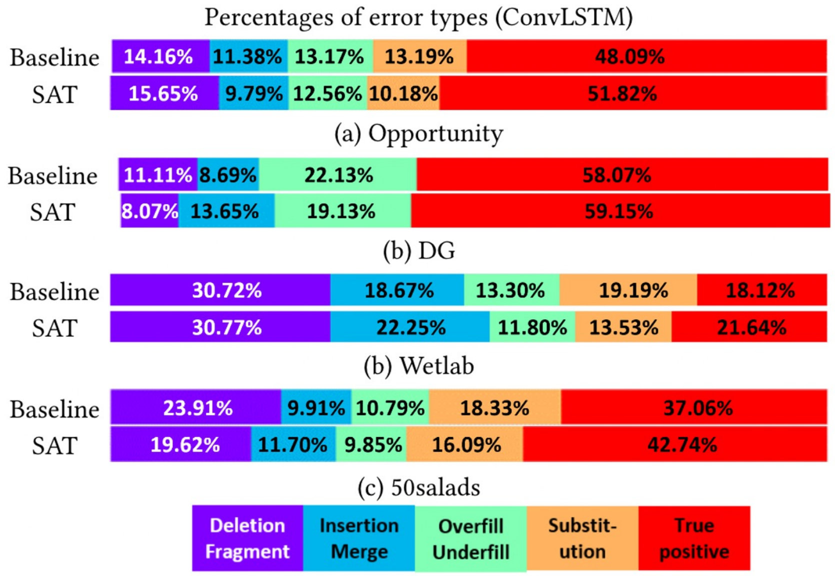


Figure 2: Smooth Activity Transition (SAT) with adaptive label jittering distributions.

Evaluation Procedure

- **Datasets:** Opportunity, Daphne Gait, Wetlab, 50salads
- **Metrics:** F1 score, Wilson score interval, misalignment analysis, …
- **Models:** Random Forest (RF), Conv, ConvLSTM
- **Labels:** baseline (remove the uncertainty), SAT-OV, SAT-TR, SAT

Results



Misalignment Analysis

	Opp	DG	Wetlab	50salads
Random Forest (RF)				
Baseline	0.367	0.286	0.168	0.467
SAT-TR	0.389	0.335	0.188	0.474
SAT-OV	0.403	0.334	0.179	0.498
SAT	0.428	0.349	0.194	0.511
Conv				
Baseline	0.643	0.703	0.315	0.536
SAT-TR	0.652	0.725	0.338	0.576
SAT-OV	0.668	0.717	0.335	0.591
SAT	0.679	0.740	0.346	0.609
ConvLSTM				
Baseline	0.672	0.734	0.341	0.564
SAT-TR	0.675	0.738	0.380	0.587
SAT-OV	0.680	0.736	0.376	0.596
SAT	0.698	0.742	0.396	0.615
Train-core-only	0.645	0.609	0.347	0.562
confidence interval	±0.009	±0.017	±0.006	±0.008
U^{OV} (p_u)	0.21s (10%)	1.52s (15%)	3.14s (10%)	0.46s (15%)

Mean F1 score

Results

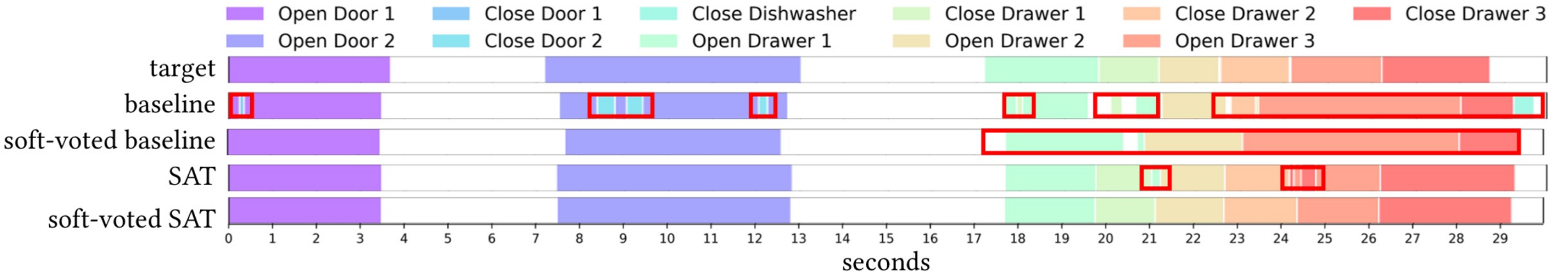
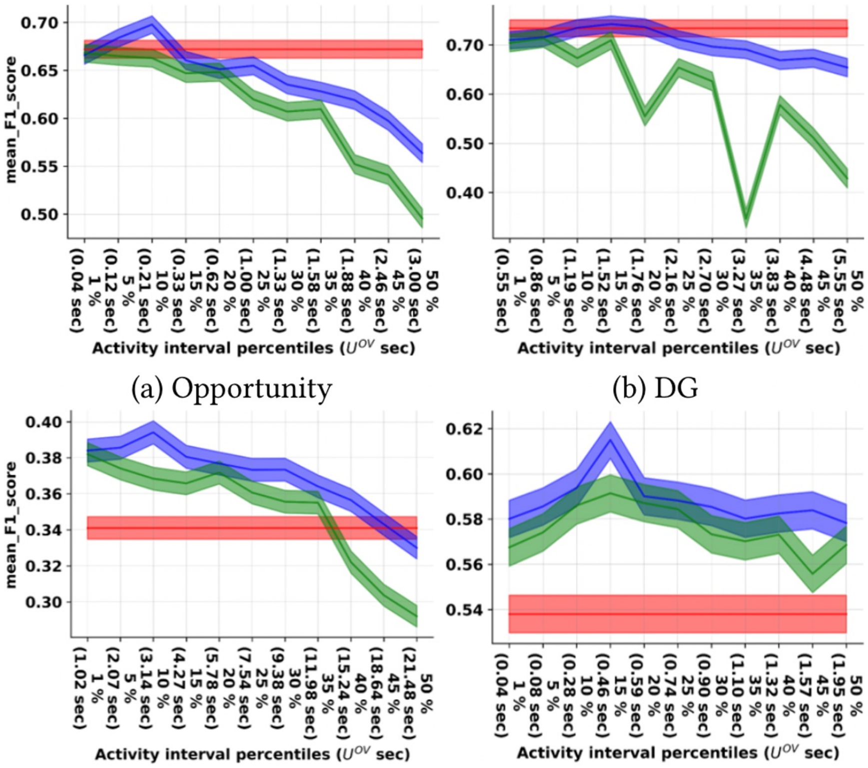


Figure 3: Importance of modeling annotation uncertainty for a sequence (30s) of complex interactions with kitchen objects (from Opportunity, ConvLSTM). Target labels (top) are compared to predictions from baseline (2nd), baseline with soft-voting (3rd), SAT (4th), and SAT with soft-voting (bottom). The significant deviations in predictions are highlighted as red boxes.

- Examining SAT compatibility with post-hoc label errors correction
- Apply soft-voting filter to the baseline/SAT labels
    - The filter removes sequence of activities in the baseline case
    - The filter do minor adjustments on the SAT case, results in a sequence nearly identical to the target.

Effects of Adaptive Label Jitter Modeling



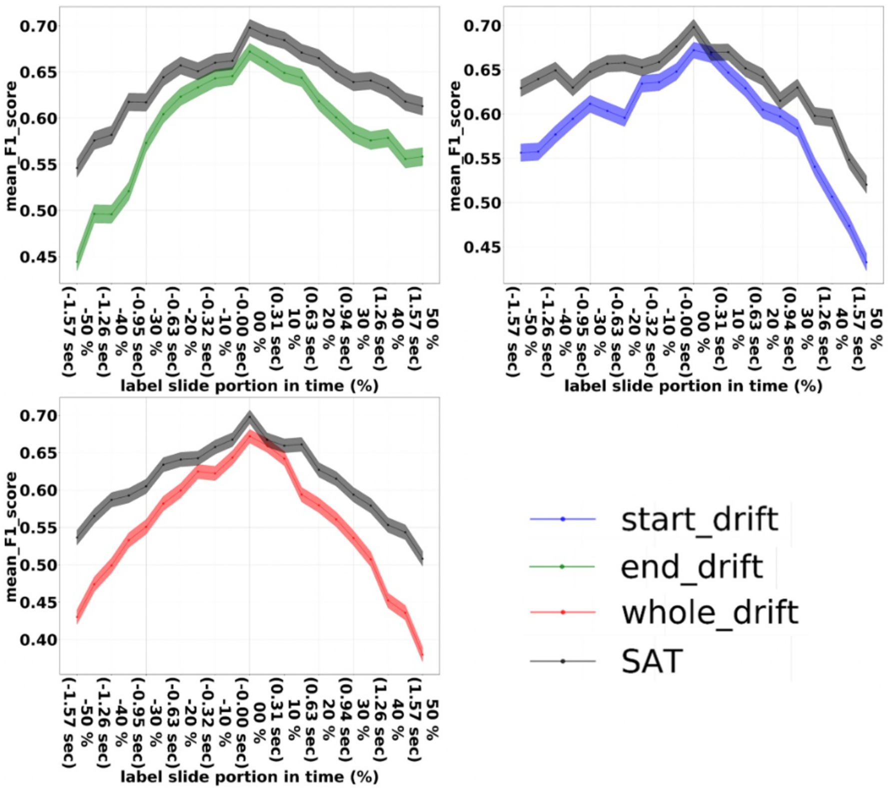
(a) Opportunity (b) DG

(c) Wetlab (d) 50salads

Compare SAT with a fixed-size label jitter (uniform smoothing):

- Uniform smoothing works for dataset with large uncertainty
- Uniform smoothing underperforms the baseline in dataset with short activities/NULL intervals

Robustness Under Label Shift



Conclusion

- Create a scheme to incorporate label uncertainty into model training
- Dynamic label jitter smoothing significantly increases model performances on various datasets
- Adaptive SAT is shown to reduce boundary errors and substitution errors
- **Future research:**
    - Does uncertainty boundary depend on activity classes?
    - Does long-term context influence annotation behaviors?