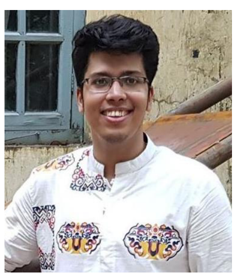
Unleashing the Power of Shared Label Structures for Human Activity Recognition

Xiyuan Zhang, Ranak Roy Chowdhury, Jiayun Zhang, Dezhi Hong,
Rajesh Gupta, Jingbo Shang
University of California, San Diego



Outline

- Motivation
- Method
- Experiments
- Conclusion

Outline

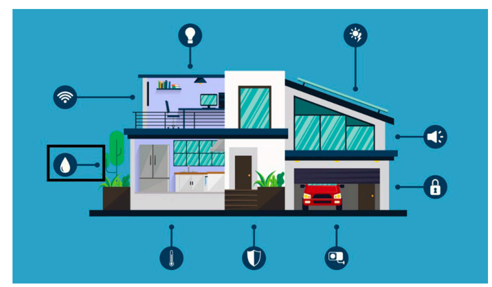
- Motivation
- Method
- Experiments
- Conclusion

Motivation

- Human Activity Recognition (HAR): identifies human activities using sensor readings from wearable devices
- Applications
    - Healthcare, motion tracking, smart home automation, etc.

Healthcare

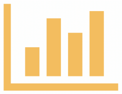
Motion tracking



Smart home automation

Motivation

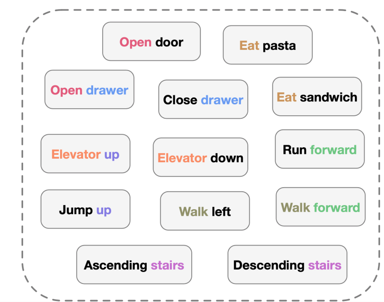
Limitation

- HAR data remain difficult to collect
  - Security or privacy concerns 
  - Certain types of human activities happen less frequently by nature 
- Existing HAR or time-series classification methods treat labels simply as integer IDs and learn their semantics purely from annotated sensor data
  - Less effective especially when labeled data are limited

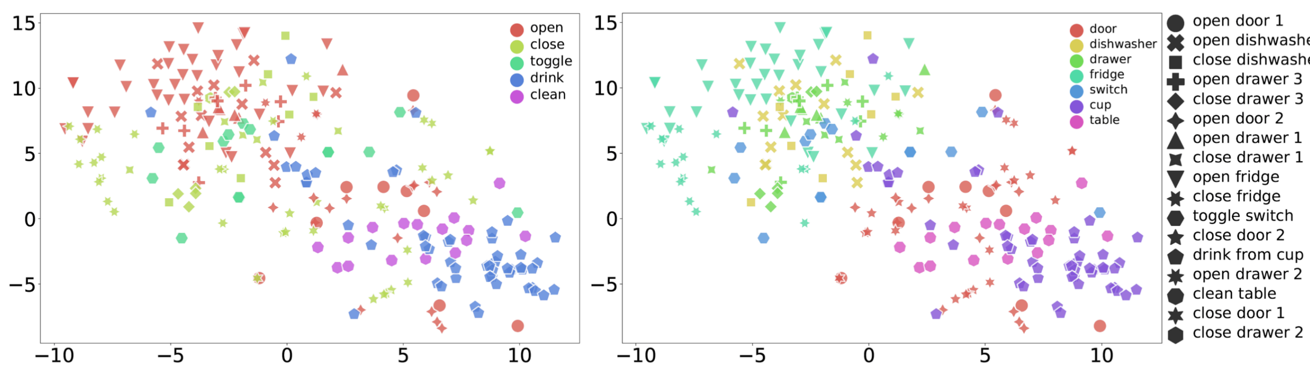
Motivation

Key Observation

- Activity names in HAR datasets often share structures that reflect the similarity between different activities
- Such mapping between input features and label names motivates the learning framework that extracts knowledge from label structures



Shared label structures in HAR data are common

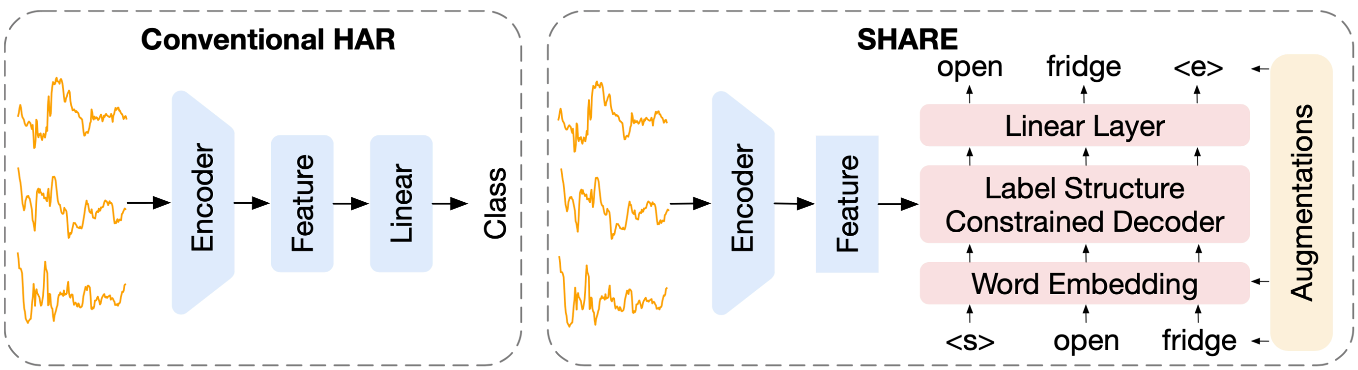


Similarity in label space translates to sensor data space

Motivation

Key Observation

- Conventional HAR: output integer class IDs as prediction results
- Our SHARE: output label name sequences, therefore preserve structures and relationships among various activities



Motivation

Key Contributions

- A more effective HAR framework by modeling label structures
  - Token-level augmentation
  - Embedding-level augmentation
  - Sequence-level augmentation

Pre-trained model enhanced augmentation

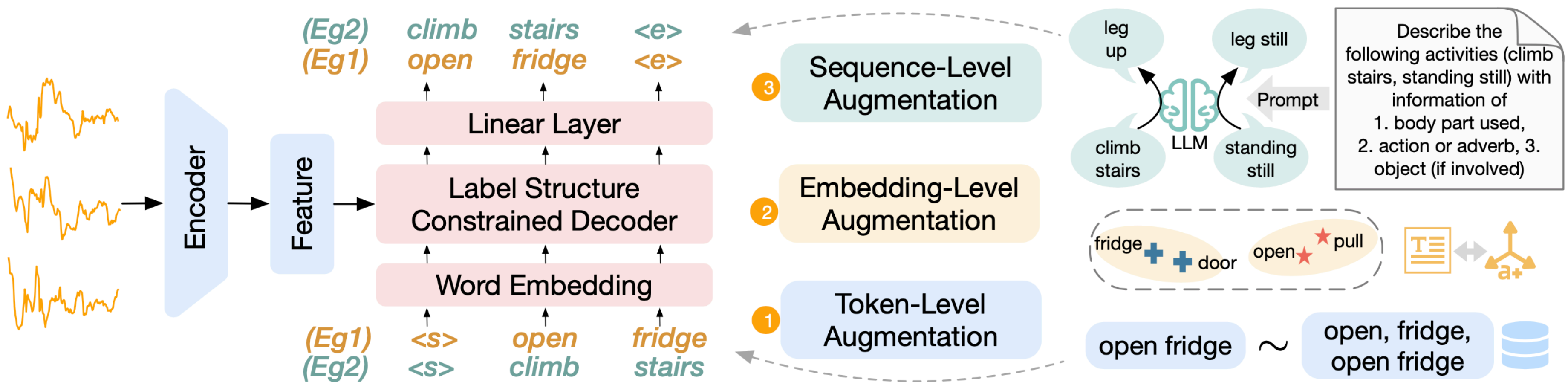
- State-of-the-art HAR performance on seven benchmark datasets
  - Especially in few-shot and label imbalance settings

Outline

- Motivation
- Method
- Experiments
- Conclusion

Method
Framework

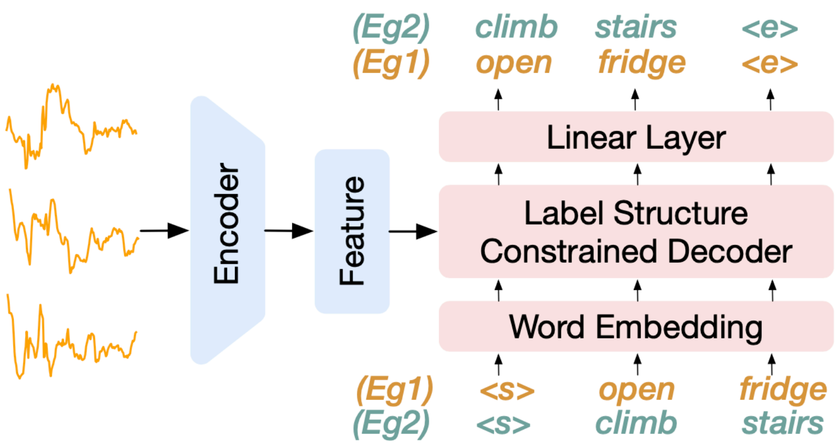
- Overall framework



Method

Encoder-Decoder

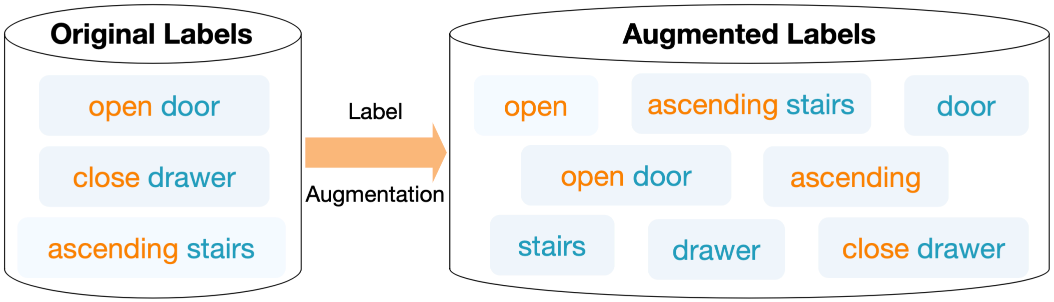
- Encoder: CNN $f_\theta : \mathcal{X} \rightarrow \mathcal{Z} \subset \mathbf{R}^d$
- Decoder: LSTM $g_\phi : \mathcal{Z} \rightarrow \mathcal{Y}$
    - Generate word sequences
    - Training: cross entropy loss
    - Inference: constrained decoding (choose the sequence with the highest probability only among valid label sequences)



Method

Augmentation

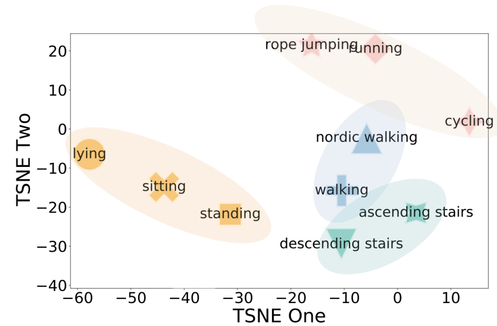
- Token-level augmentation
  - Better learn semantics of each token in the label sequence
  - Randomly choose meaningful single words from the original label sequence as the new labels, e.g., “open” from “open door”



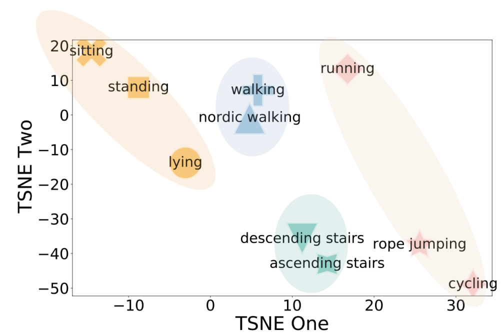
Method

Augmentation

- Embedding-level augmentation
  - Capture implicit semantic structures that are not explicitly presented as shared label names (e.g., “walk”, “run”)
  - Capture semantics by word embeddings from pre-trained models (ImageBind)
  - Pre-trained word embeddings to initialize decoder’s word embedding layer



Data

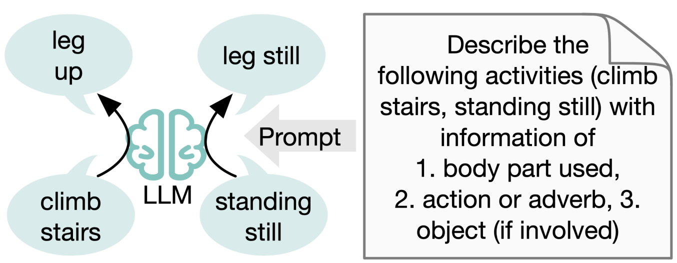


ImageBind embeddings

Method

Augmentation

- Sequence-level augmentation
  - Some datasets do not have or rarely have shared label names
  - Use large language models to generate label names with shared tokens



Original Label Names	Generated Label Names
standing still	leg still
sitting and relaxing	buttocks still
lying down	back down
walking	leg walk
climbing stairs	leg up
waist bends forward	back forward
frontals elevation of arms	arm up
knees bending (crouching)	leg forward
cycling	leg cycle
jogging	leg jog
running	leg jog fast
jump front and back	leg jump

Example generated label names on Mhealth data

Outline

- Motivation
- Method
- Experiments
- Conclusion

Experiments

Datasets

- HAR benchmark datasets

Dataset	Train	Test	Window Size	Channel	Class Num	An Example Subset of Shared Label Names
Opportunity [38]	2891	235	150	45	17	open door, open drawer, close drawer, open fridge, open dishwasher
PAMAP2 [37]	14438	2380	512	27	12	ascending stairs, descending stairs, walking, nordic walking
UCI-HAR [1]	7352	2947	128	9	6	walk, walk upstairs, walk downstairs
USCHAD [55]	17576	9769	100	6	12	run forward, walk forward, elevator up, elevator down, jump up
WISDM [46]	12406	3045	200	6	18	eating soup, eating pasta, kicking soccer ball, playing tennis ball
Harth [28]	14166	3588	300	6	12	sitting, standing, cycling sitting, cycling standing, cycling sitting inactive

Experiments

Main Results

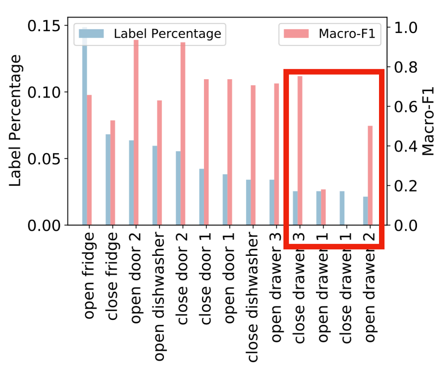
- State-of-the-art performance compared with HAR and time-series classification baselines

Datasets	Metrics	DeepConvLSTM [33]	XGBoost [7]	MA-CNN [36]	HHAR-net [14]	TST [52]	TARNet [10]	Rocket [12]	THAT [24]	SHARE
Opp	Accuracy	0.746±0.049	0.688±0.017	0.549±0.029	0.753±0.027	0.784±0.018	0.789±0.024	0.811±0.008	0.803±0.012	**0.849±0.015**
	Macro-F1	0.634±0.036	0.547±0.011	0.416±0.036	0.620±0.021	0.668±0.023	0.669±0.034	0.670±0.016	0.691±0.015	**0.766±0.013**
PAMAP2	Accuracy	0.891±0.012	0.939±0.003	0.926±0.011	0.885±0.031	0.922±0.037	0.931±0.011	0.928±0.008	0.943±0.005	**0.960±0.002**
	Macro-F1	0.884±0.018	0.939±0.007	0.925±0.012	0.893±0.031	0.925±0.039	0.935±0.010	0.934±0.008	0.949±0.005	**0.965±0.002**
UCI-HAR	Accuracy	0.900±0.016	0.907±0.003	0.921±0.025	0.926±0.005	0.926±0.005	0.904±0.011	0.939±0.002	0.906±0.007	**0.960±0.002**
	Macro-F1	0.899±0.016	0.906±0.003	0.921±0.024	0.926±0.005	0.925±0.006	0.904±0.011	0.942±0.002	0.909±0.006	**0.959±0.002**
USCHAD	Accuracy	0.574±0.016	0.571±0.007	0.543±0.044	0.524±0.011	0.641±0.028	0.564±0.037	0.580±0.005	0.643±0.015	**0.674±0.041**
	Macro-F1	0.557±0.015	0.573±0.006	0.520±0.047	0.523±0.009	0.594±0.023	0.533±0.021	0.601±0.007	0.619±0.012	**0.627±0.027**
WISDM	Accuracy	0.689±0.014	0.668±0.005	0.634±0.059	0.566±0.016	0.715±0.003	0.733±0.011	0.643±0.007	0.774±0.005	**0.794±0.003**
	Macro-F1	0.685±0.013	0.662±0.006	0.631±0.060	0.538±0.012	0.710±0.004	0.737±0.010	0.767±0.004	0.634±0.005	**0.790±0.004**
Harth	Accuracy	0.979±0.006	0.977±0.001	0.973±0.016	0.981±0.001	0.974±0.005	0.962±0.009	0.897±0.003	0.960±0.016	**0.983±0.007**
	Macro-F1	0.578±0.032	0.522±0.003	0.538±0.025	0.515±0.049	0.501±0.031	0.481±0.031	0.472±0.019	0.485±0.025	**0.593±0.020**

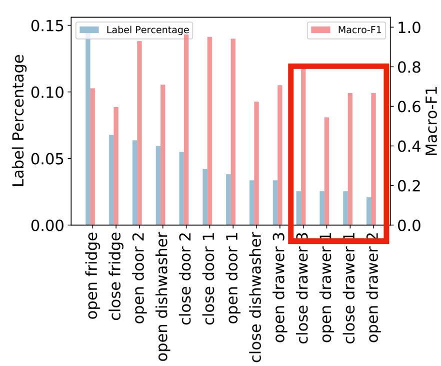
Experiments

Few-Shot Learning

- More significant improvement in few-shot settings

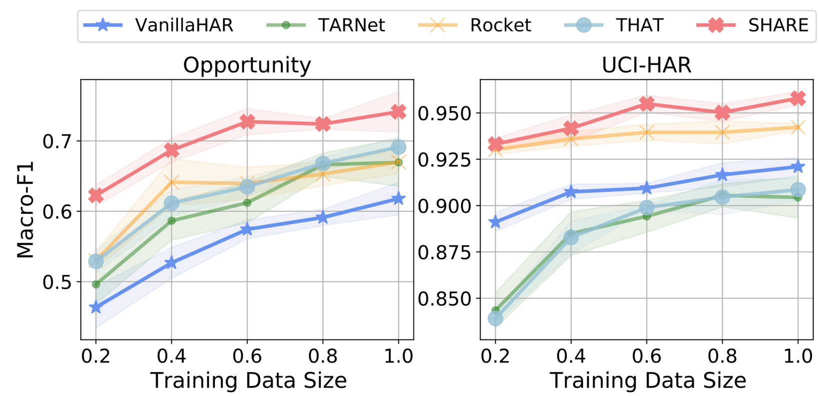


Integer-ID based HAR



SHARE

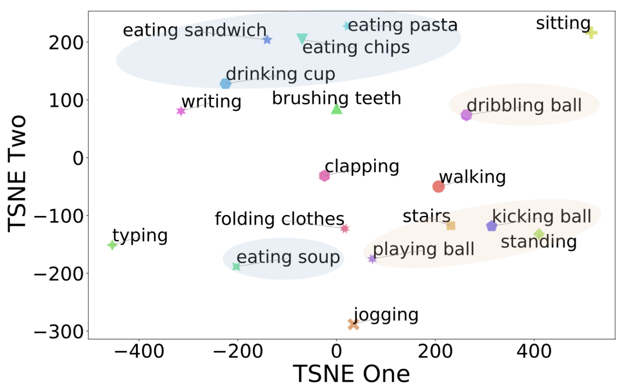
Long-tail distribution



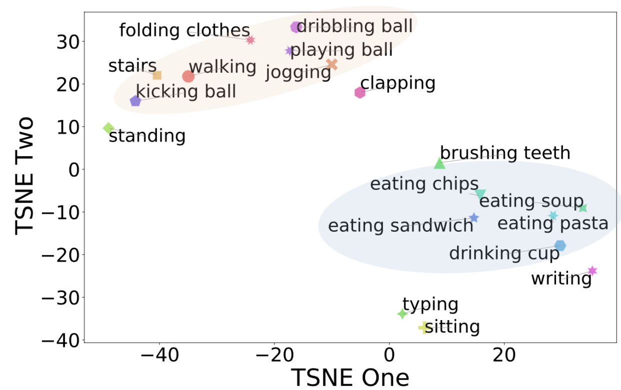
Reduced training data

Experiments
Case Study

- T-SNE visualization on feature space
- SHARE better preserves semantics in the feature space



Integer-ID based HAR



SHARE

Experiments

Complexity Analysis

- SHARE is more time and memory efficient compared with state-of-the-art deep learning models

Model	# of Params	Model Size	Avg Running Time Per Batch
TST	1.195M	4.786MB	0.014s
TARNet	0.310M	2.465MB	0.016s
THAT	3.207M	12.828MB	0.018s
SHARE	**0.219M**	**0.878MB**	**0.003s**

Outline

- Motivation
- Method
- Experiments
- Conclusion

Conclusion

- A new HAR framework, SHARE
  - Models label semantic structures
  - Three label augmentations leveraging large language models for enhanced semantics modeling
- State-of-the-art HAR performance on seven HAR benchmark datasets
  - Especially in few-shot and label imbalance settings

Thank you!

Contact: xiyuanzh@ucsd.edu

Code Release: https://github.com/xiyuanzh/SHARE

Python Package: https://pypi.org/project/semantichar/