NTNU

Norwegian University of Science and Technology

Adversarial Unsupervised Representation Learning for Activity Time-Series

Presented at AAAI-19

https://arxiv.org/abs/1811.06847

Motivation

Huge amount of data from wearable devices, monitoring user's physical activity and sleep

Why not use these to assist healthcare professionals?

Problem description

- Few subject have clinical data AND activity data

- A lot of activity data, a supervised approach would result in most of this being unused

- Wearables' signals depends on the person wearing them as well as their environment

Related work - topic 1

Human activity for health informatics - use of activity data

- human activity recognition
- manual diagnostication of sleep-disorders
- quantifying sleep quality using DL
- monitoring human behavioural patterns

This paper propose task-agnostic models instead of plain supervised learning

Related work - topic 2

Representation learning - constructing a space that is discriminative for downstream task

- Better network convergence by adding unsupervised pre-trained vectors that encode mutual information between the input features
- Distributed bag-of-words architectures that predicts the context of the structure

This paper integrates a DBOW architecture in an adversarial setting to capture local patterns

Related work - topic 3

Time series analysis literature

- use of pair-wise similarity concept to perform classification and clustering tasks
- SAX and BOSS convert time-series to a symbolic sequence - both are supervised
- HCTSA feature extraction - unsupervised

Embeddings learned by this paper's model can be used to initialize supervised learning models

Methods

1) Demonstrating the model's behaviour for granularities that are intuitive to humans

2) Dealing with local patterns within a time segment

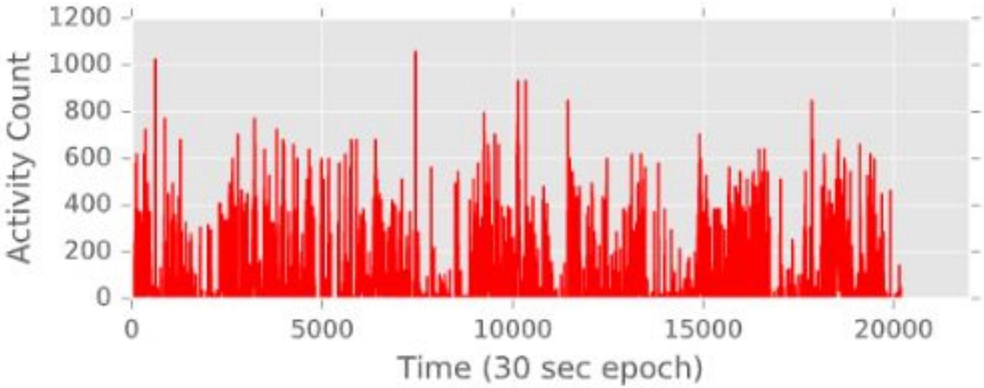


Figure 1: Activity time-series for a subject over a week.

Methods

3) Dealing with global patterns between time segments

4) Make sure that subject-specific noise and environment does not affect the activity data

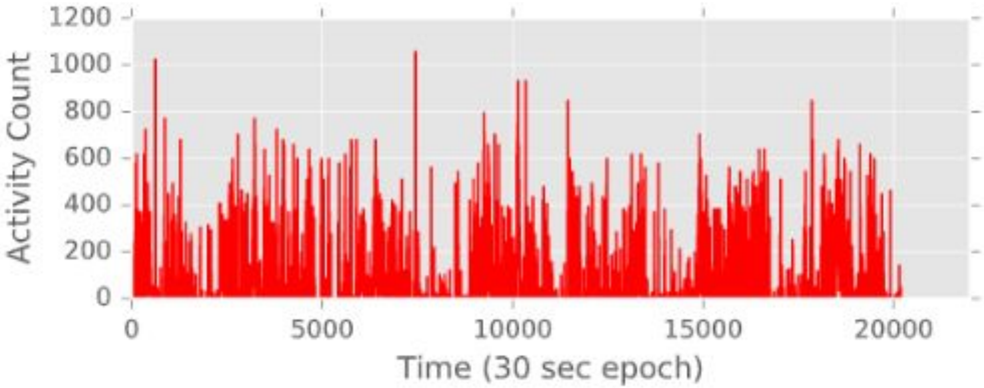


Figure 1: Activity time-series for a subject over a week.

Experimental Settings - datasets

SOL

- activity data and clinical data for 1887 subjects

MESA

- activity data for 2237 subjects

Together these datasets simulate the common scenario for semi-supervised learning

Experimental Settings - tasks

Evaluating the effectiveness of the learned embeddings

Binary classification tasks:

- Sleep Apnea
- Hypertension

Three-class classification tasks:

- Diabetes
- Insomnia

NTNU

Results

	Method
Supervised	Majority
	Random
	SAX-VSM
	BOSS
	BOSSVS
	Task-specific
Unsupervised	sample2vec
	hour2vec
	hour2vec+Reg
	day2vec
	day2vec+Reg
	week2vec
	HCTSA
	LSTM
	day2vec+Reg+O
	day2vec+Reg+O+A

Capturing local patterns:

Segment-Specific Loss

Ordinal Regression Loss (O)

Capturing global patterns:

Neighbour Context Loss

Smoothing Loss (R)

Subject invariance:

Adverseral Loss (A)

Results

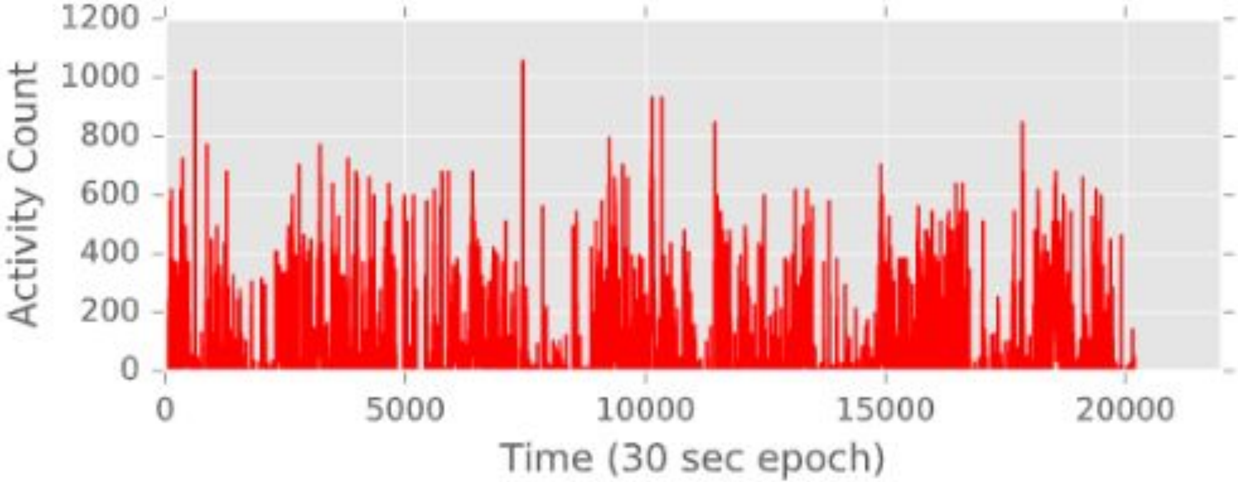


Figure 1: Activity time-series for a subject over a week.

Results - adversarial loss

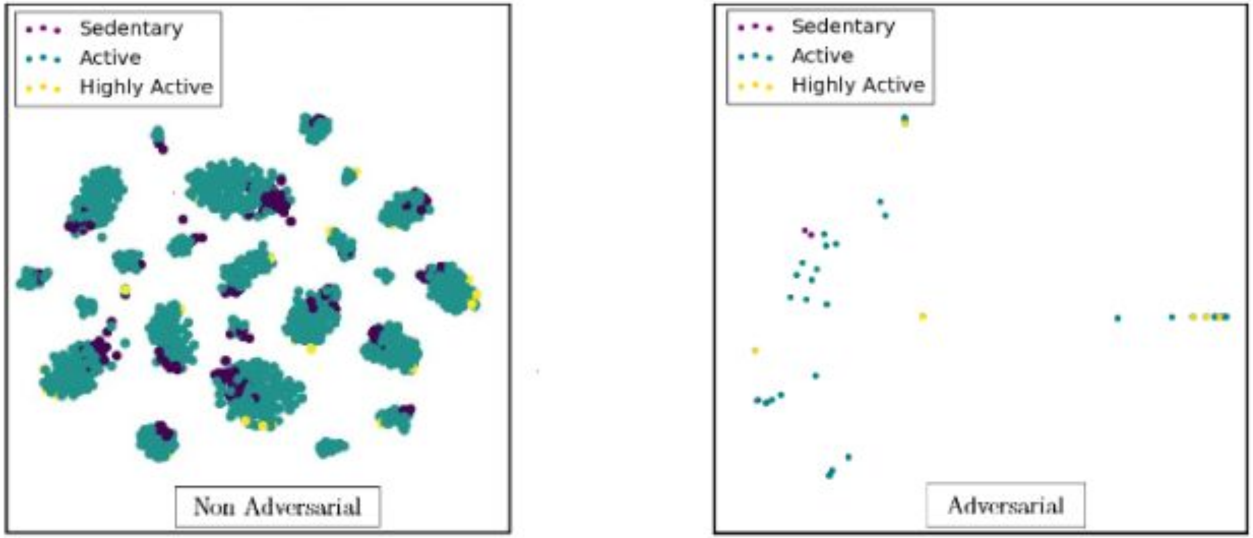
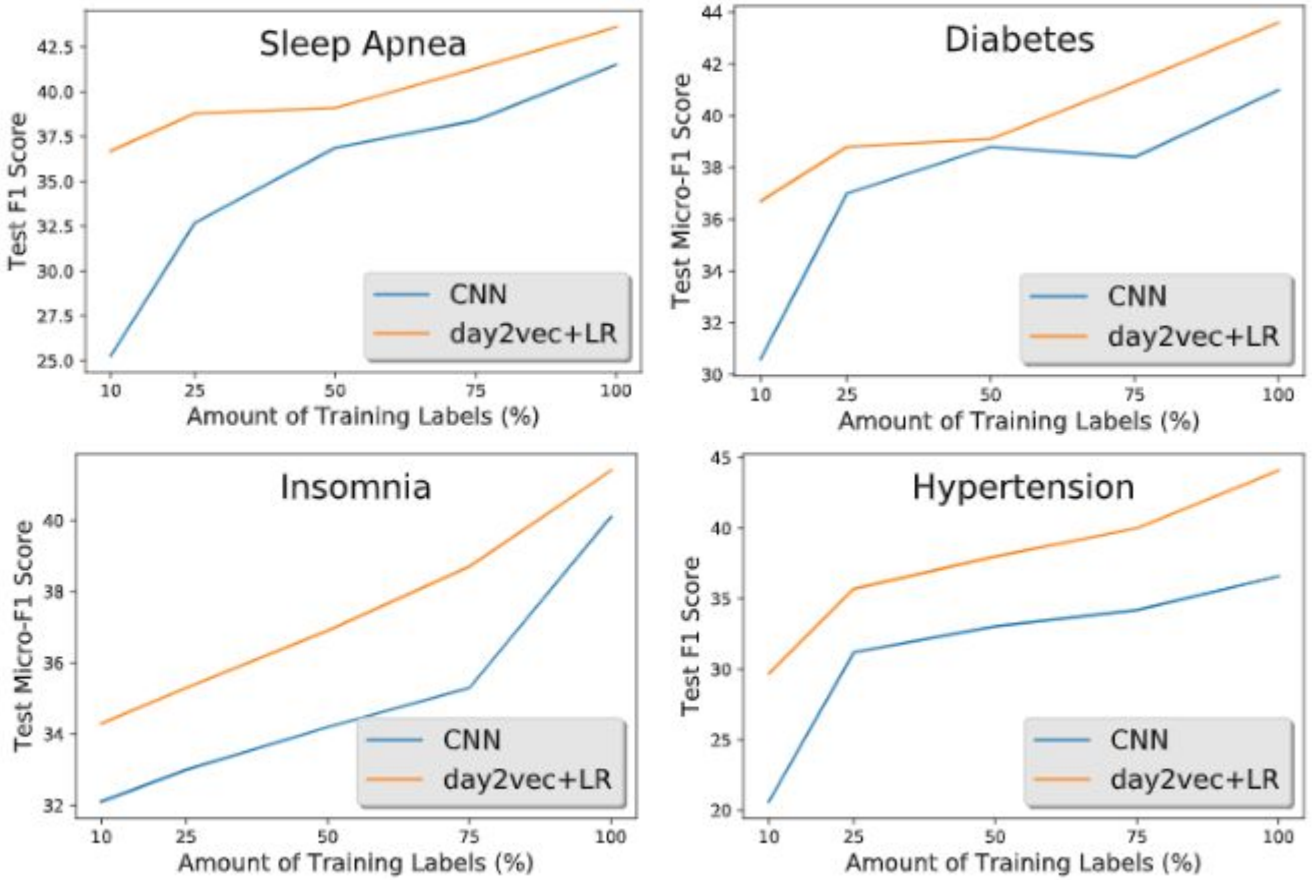


Figure 4: t-SNE visualization of subjects with regularized `day2vec` on the left and adversarial `day2vec` on the right, for all the subjects with respect to their level of activity.

Results - compared to supervised



Conclusions

- Day-level granularity for activity time-series are shown to be better (than other human-intuitive granularities)
- Unsupervised representation learning technique that works well with a simple classifier
- Generalizes embeddings
- Can complement supervised learning by initialization

Thank you