Ayala Raanan, ID: 040474934, Gil Levy, ID: 029378577

Project Proposal

Anchor Paper: AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition

Paper Summary

Domain - The paper deals with the topic of Human Activity Recognition (HAR). The source of the data is often multiple sensors, positioned on a human body and reporting the motion of the participant while performing various activities. The sensors used often include IMUs with accelerometers, gyroscopes and magnetometers. The goal of HAR models is to classify a time window of the data to the right activity. HAR domain dealing with videos is not discussed.

Related work – previous work done on the subject includes different approaches for processing the data, with emphasis on how best to fuse continuous data coming from different sensors. Earlier approaches used engineered features with classical SVM/RF models. More advanced models incorporated deep learning, mainly using CNNs for feature extraction and additional RNNs for temporal analysis.

Data processing – The conventional paradigm tries to classify an activity within a defined window of time. Each window is split into shorter non-overlapping segments called time-steps. First, data from all sensors within a time-step is processed for an intermediate representation of the time-step. Then, additional processing considers results from all time-steps to produce a prediction. There are numerous variations on the subject. State-of-the-art model prior to this paper (Deepsense, Yao et al., 2017) used CNNs as subnets for feature extraction from each sensor within each time-step. An additional CNN was then applied for fusing different sensors together. Finally, GRU layers combined data from all time-steps for a downstream task representation.

Datasets: Different types of labeled activities with number of participant, activities, and sensor types as below:

Name	Subject	S. Rate	Activity	Sample	Sensor
Heterogeneous	9	100 Hz	6	43,930,257	A, G
Skoda	1	98 Hz	11	22,000	A
PAMAP2	9	100 Hz	12	2,844,868	A, G, M

AttnSense main steps:

- Preprocessing includes
  - Segmentation to windows and splitting to equal size time-steps
  - Converting time-step data from time domain to frequency domain using FFT
  - Data augmentation by noise addition to raw data
- Per-sensor CNN subnet – data from each sensor is fed into its individual CNN for feature extraction
- Output representations from all sensors within a time-step are combined with attention weights for a single vector representation of the time-step.
- All time-steps within the classification window are processed by a double layer GRU.
- GRU outputs for each time-step are combined using another (temporal) attention layer
- A final linear-softmax head is used for classification

Paper main objectives:

- This paper was one of the first to suggest using attention to better capture spatial and temporal dependencies in the data:

    - *Spatial dependencies* – Not all sensors contribute equally to the classification of activities. Our paper uses an attention mechanism to combine data from different sensors. The model learns what weight to give each sensor for the sake of the classification task.
    - *Temporal dependencies* – Not all time-steps within a time window contribute equally. Some parts of the activity profile may be more salient than others. The paper uses another attention mechanism applied to the GRU time-step output representations.
- Using Attention mechanisms greatly improves interpretability of the model by letting us see which sensors are more important (which helps in designing real commercial systems) and which time-steps are more important.

Results – We wish to focus on the PAMAP2 dataset and reproduce F1-score of about 0.89 on this dataset. The Skoda dataset is also available if time will allow.

Innovation

We propose to explore 2 types of improvements to the existing model for both data processing and model:

1. *Data augmentation*: It seems natural that the signals from some of the sensors used are susepetable to rotations. Especialy since they can be installed differently for each person. An accelerometer for example would see gravity at a different angle. For the best of our knowledge, the data augmented for HAR models only includes addition of normally distributed noise. We propse to augment data by applying random rotation on the raw data as a first step. Rotation will be applied to all correlated axes of a sensor together. We hope that such augmentation would make the model more robust to naturally occuring rotation. We can compare this type of improvement to noise addition.

2. *Model*: The architecture of the existing model offers only one attention vector to be learnt for temporal dependencies. This seems very limiting for learning activity-specific temporal dependencies. To allow such specificity, we propose to explore the use of multihead temporal attention.

We will use F1-score on the same dataset to evaluate our innovation.

Timeline and milestones

Milestone Description	**Due-date**
Anchor Paper	
Dataset analysis and statistics (at least one dataset)	10/July
Attnsense model construction and training (Not including augmented preprocerssing, various convulation options, sequecne length etc…)	23/July
Full reconstrunction of F1-score: Including preprocessing features, optimizing model architecture parameters, exploring the contribution of various model features (Especially FFT & attension layers) to the final F1-score Note: At there is no code available for this paper and the paper lacks many important details regarding exact architecture and implementation we expect challenges in achieiving simillar results. Anchor report submission	31/July
Innovative part	
Implement data augmentation by rotation – compare to random noise addition	13/Aug
Implement multihead attention – compare to base model Final report submission	28/Aug