Improving Human Activity Recognition with Domain Knowledge Composition

Massinissa Hamidi

hamidi@lipn.univ-paris13.fr
hamidimassiniss

April 18th, 2019


MODELS.
MODELS EVERYWHERE
imgflip.com

Context & motivation

- We are increasingly surrounded by things enabled with perception capabilities;
- Autonomous vehicles, industry 4.0, smart homes, smart cities;
- Monitoring of industrial infrastructures: more than 1.2 trillion observations generated each year [XHC+14];
- Autonomous vehicle: 40 Terabytes of data generated every 8 hours of driving;
- Monitoring of pollution using a car fleet in a city or region [LKZ+18];

Context & motivation

- **Diversity of modalities**, *eg.* temperature, pressure, sound, *etc.*;
- **Diversity of sensors specs**, *eg.* precision, reactivity, functioning conditions, *etc.*;
- **Diversity of topologies**: sensors located at different positions of the space or the scene of interest;
- **Nature of deployments** which are continuously evolving either in terms of components or topologies;

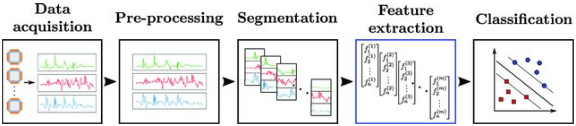


Figure: Traditional learning pipeline [BBS14, LSN+18]

Problematics

1. **Robustness of learning processes**;
2. **Will current approaches be able to scale-up?**;
3. **Learning processes in goal-independent and evolving deployements**;

Approach

Rather than considering the large mass of generated data, our approach consists in handling data sources in an intelligent way.

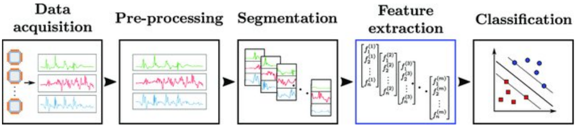
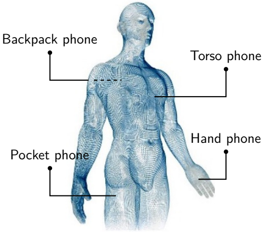


Figure: Traditional learning pipeline [BBS14, LSN+18]

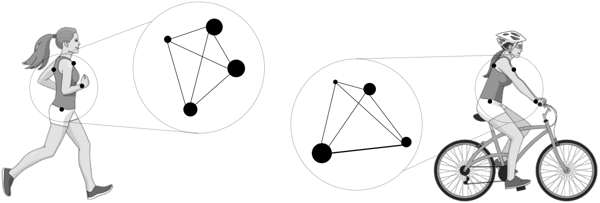
This could be done via a **modeling** effort of the different steps of learning pipelines and a formalization of **models composition**.

The importance of infrastructure

Application to the Sussex-Huawei Locomotion (SHL) dataset [GCW+18]



(a) Topology of data sources in the SHL dataset



(b) Example of the relative importance of each position and the level of interactions between modalities for the recognition of running and bicycling activities.

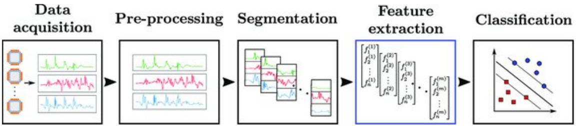
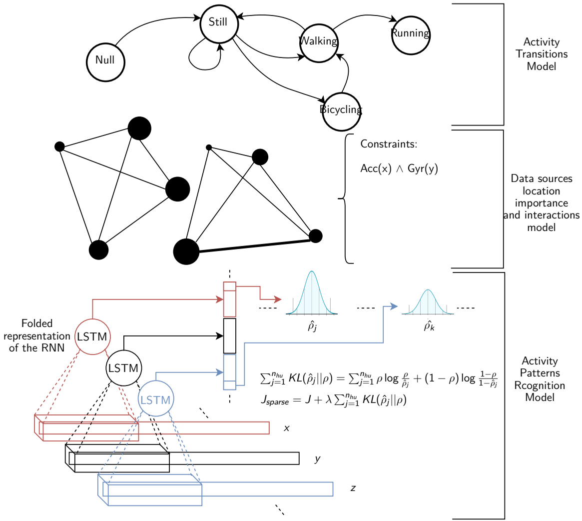


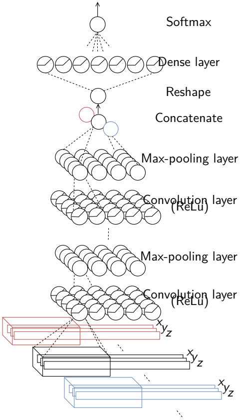
Figure: Traditional learning pipeline [BBS14, LSN+18]

Composition of models

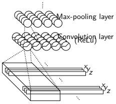


Modeling data sources interactions and importance

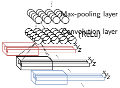
Convolutional modes, Bayesian optimization of hyperparameters, and functional analysis of variance fANOVA.



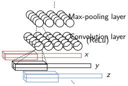
Convolutional neural architecture



modalities grouped



Split modalities



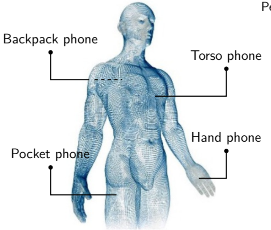
Split channels

Param.	low	high	prior
l_r	0.001	0.1	log
$ks_{1,mod}$	9	15	-
$ks_{2,mod}$	9	15	-
$ks_{3,mod}$	9	12	-
$n_{f,mod}$	16	28	-
s_{mod}	0.5	0.6	log
p_d	0.1	0.5	log
n_u	64	2048	-

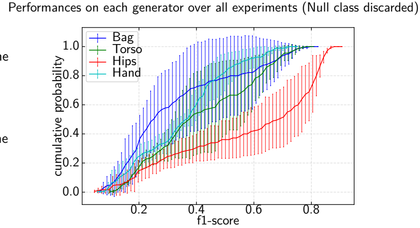
Summary of the different hyperparameters assessed during Bayesian optimization process along with their respective bounds.

Hyperparameter/Architecture space exploration

- Large-scale experimentation (more than 5k models);



(e)



(f)

Figure: Cumulative distribution of the performances of the models trained on each body-position. Bayesian optimization results averaged across users and positions.

Influence of data source location

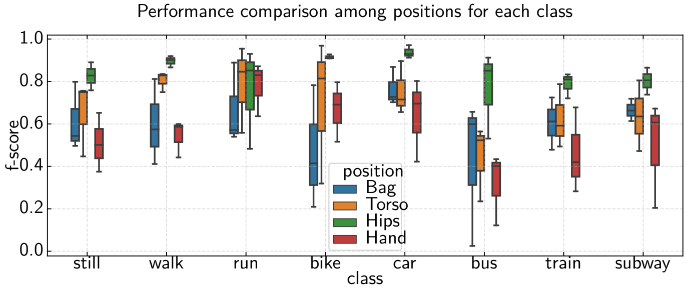


Figure: Contribution of the data sources to the overall recognition performances of each locomotion mode. Data sources are grouped by their respective positions. The analysis combines the entire configurations obtained via the Bayesian optimization runs.

These results corroborate findings in [FSF99, MHS01] on the suitability of the hips location for recognizing different kinds of human activities.

Emerging modalities and interactions

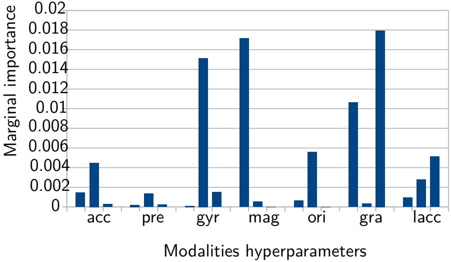


Figure: Individual marginal importance of the kernel size hyperparameters controlling the impact of each modality.

Hyperparam.	Pairwise marginal ($\times 10^{-4}$)
$(ks_{gyr,2},\ ks_{gra,2})$	9.2778
$(ks_{mag,1},\ ks_{ori,2})$	7.0166
$(ks_{gyr,2},\ ks_{ori,2})$	5.5122
$(ks_{acc,1},\ ks_{mag,1})$	4.0382
$(ks_{pre,1},\ ks_{gyr,3})$	2.3154
$(ks_{gyr,3},\ ks_{mag,1})$	2.2472
$(ks_{mag,1},\ ks_{ori,1})$	2.1216
$(ks_{pre,3},\ ks_{gyr,2})$	1.76305

Table: Most important pairwise marginals of the kernel size hyperparameter computed via the fANOVA framework for each data source. Results are grouped by their respective modalities.

Opening

- Proving adequation between influence of data sources and hyperparameters tuning;
- Non-convexity and convergence of the learning process?;

Q&A

Slides available from:

http://lipn.univ-paris13.fr/~hamidi/junior-seminar-2019.pdf

References I

- Andreas Bulling, Ulf Blanke, and Bernt Schiele, *A tutorial on human activity recognition using body-worn inertial sensors*, ACM Computing Surveys (CSUR) (2014).
- F Foerster, M Smeja, and J Fahrenberg, *Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring*, Computers in Human Behavior **15** (1999), no. 5, 571–583.
- Hristijan Gjoreski, Mathias Ciliberto, Li Wang, Francisco Javier Ordonez Morales, Sami Mekki, Stefan Valentin, and Daniel Roggen, *The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices*, IEEE Access (2018).
- Yuxuan Liang, Songyu Ke, Junbo Zhang, Xiuwen Yi, and Yu Zheng, *Geoman: Multi-level attention networks for geo-sensory time series prediction.*, IJCAI, 2018, pp. 3428–3434.

References II

Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek, *Comparison of feature learning methods for human activity recognition using wearable sensors*, Sensors **18** (2018), no. 2, 679.

Jani Mantyjarvi, Johan Himberg, and Tapio Seppanen, *Recognizing human motion with multiple acceleration sensors*, Systems, Man, and Cybernetics, 2001 IEEE International Conference on, vol. 2, IEEE, 2001, pp. 747–752.

Xiaomin Xu, Sheng Huang, Yaoliang Chen, Kevin Browny, Inge Halilovicy, and Wei Lu, *Tsaaas: Time series analytics as a service on iot*, 2014 IEEE International Conference On Web Services (ICWS), IEEE, 2014, pp. 249–256.