RAIS

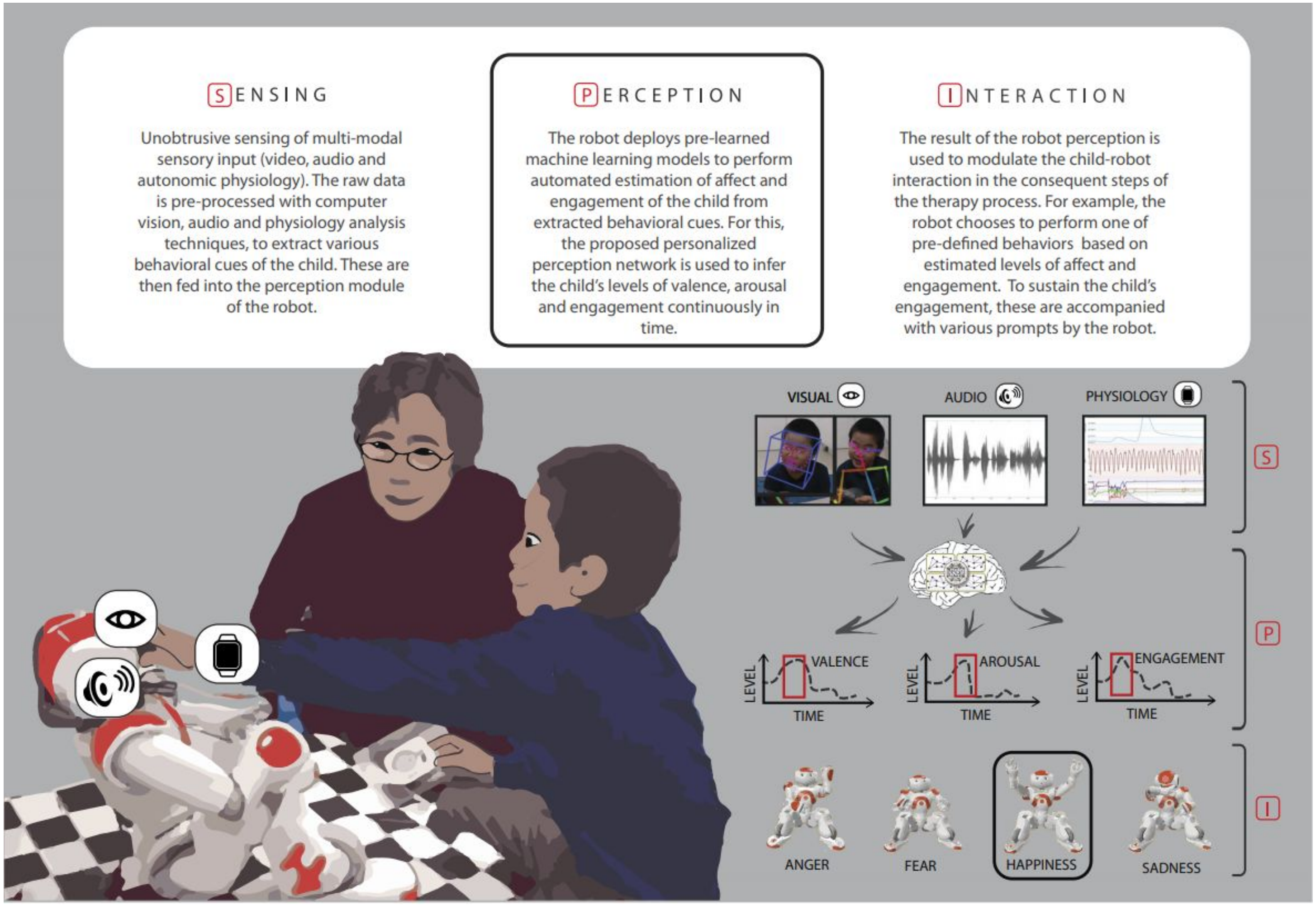
Real-time Analytics for Internet of Sports

Marie Curie European Training Network

One-size-does-NOT-fit-all: Personalized Machine Learning

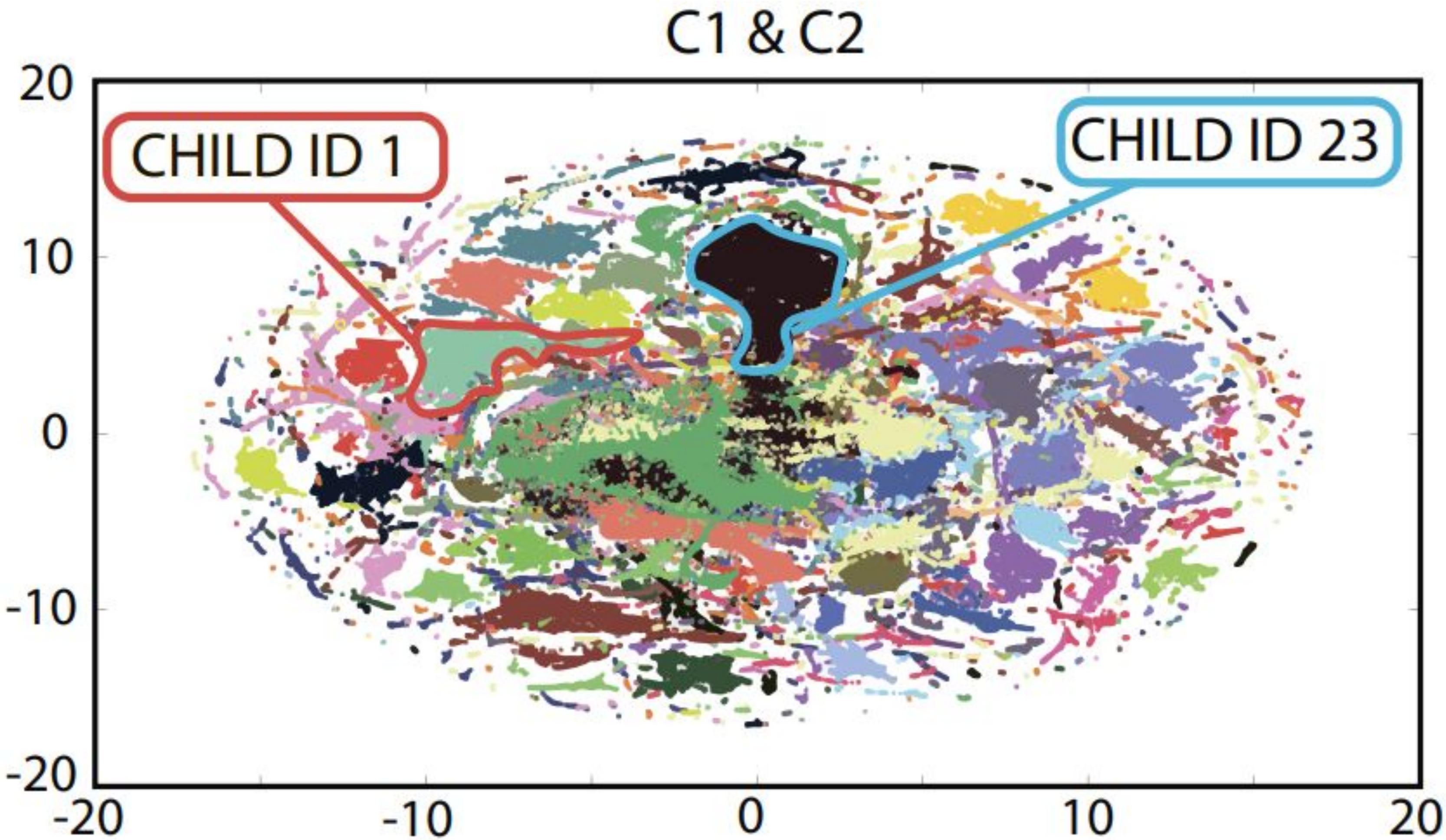
Yfantidou Sofia, Aristotle University of Thessaloniki (AUTH)

Let's start with a story...



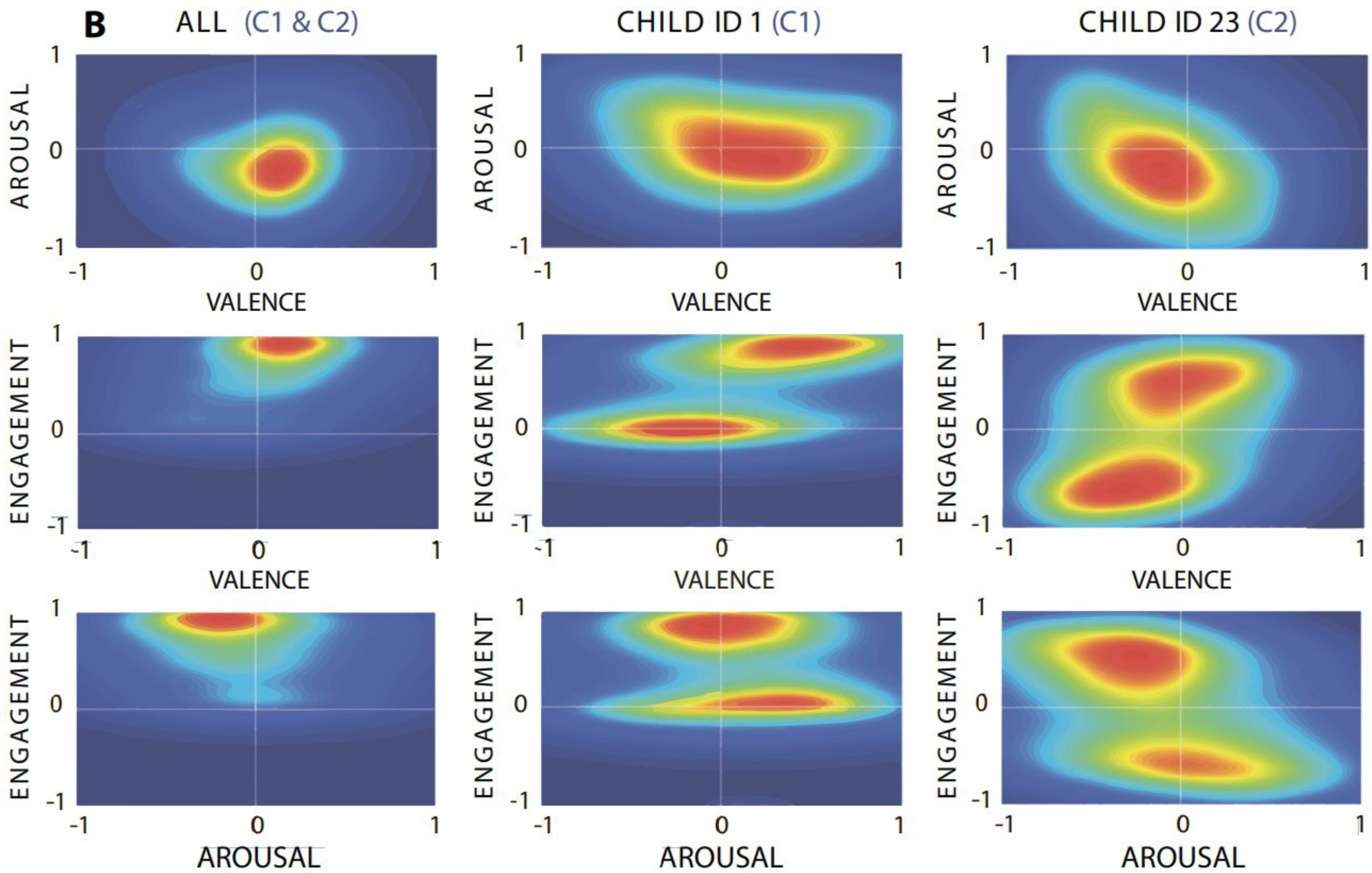
Overview of the key stages (sensing, perception, and interaction) during robot-assisted autism therapy (Rudovic et al., 2018)

How different are we?



Clustering of the children from C1 (Japan) and C2 (Serbia) using the t-SNE, an unsupervised dimensionality reduction technique, applied to the auto-encoded features (Rudovic et al., 2018)

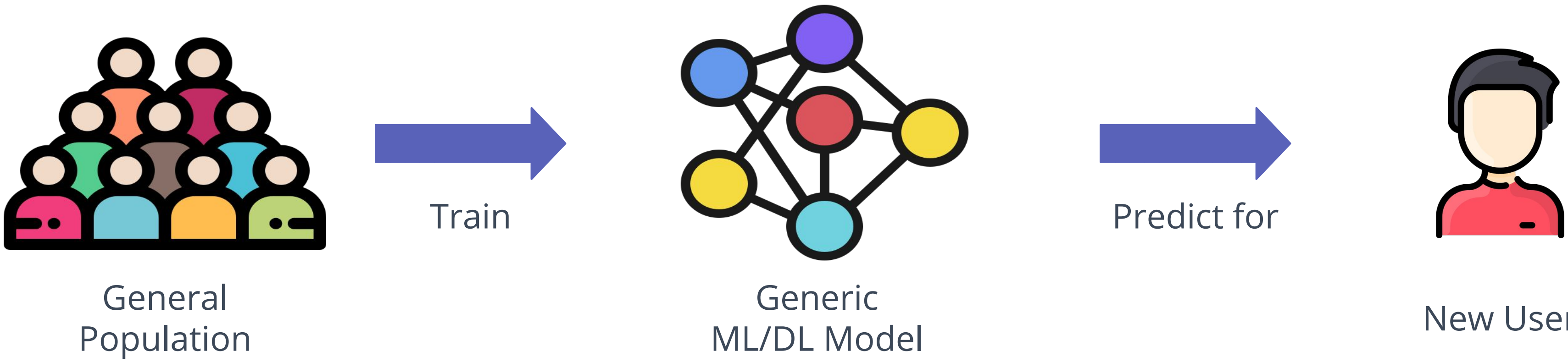
How different are we?



The heat maps of the joint distributions for the valence, arousal, and engagement levels, coded by human experts (Rudovic et al., 2018).

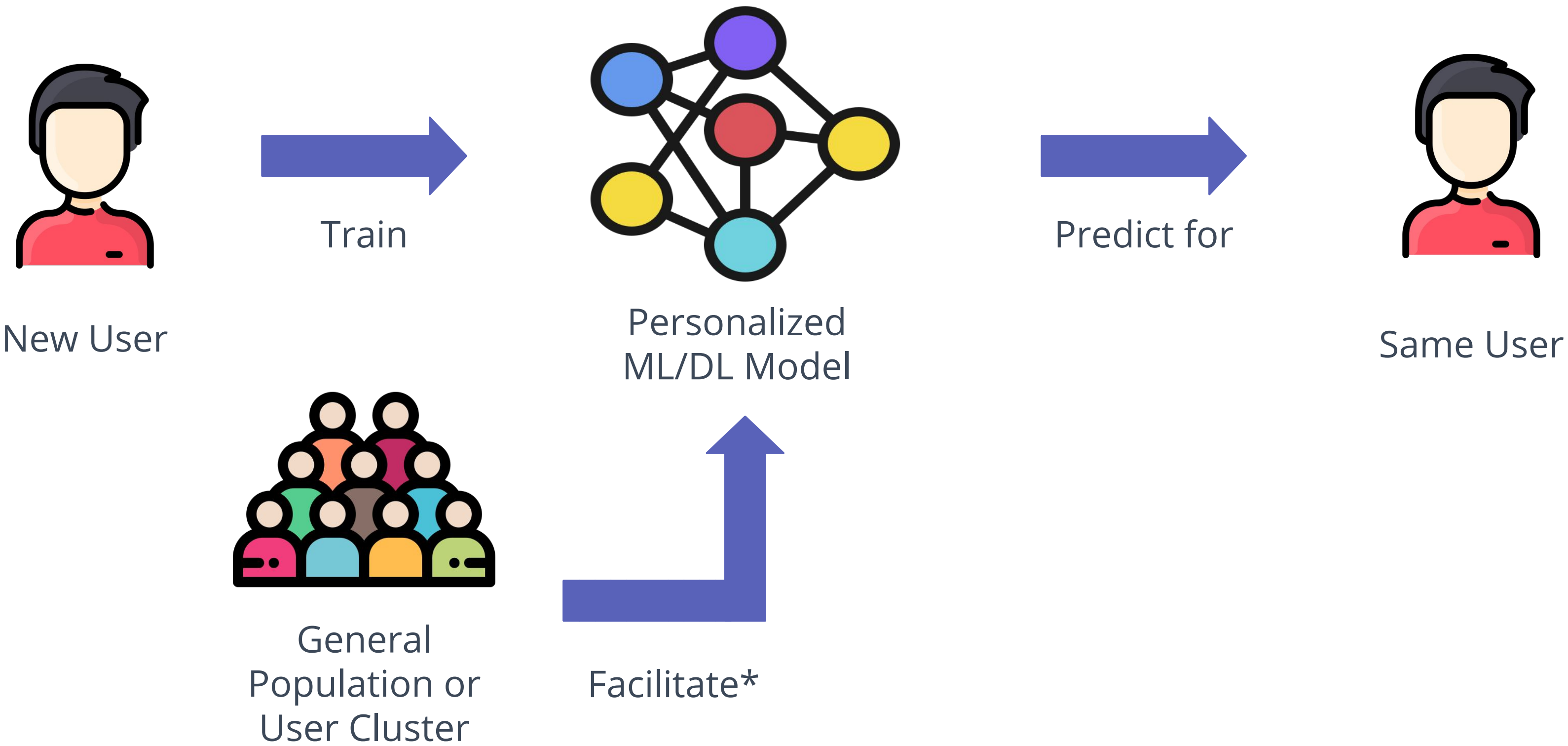
RAIS

The Problem with Traditional ML



- Generic models are tuned to an average target population
- “Good” performance doesn’t necessarily translate to each individual
- Acceptable in certain domains, but what about health and well-being?

The Solution: Personalized ML



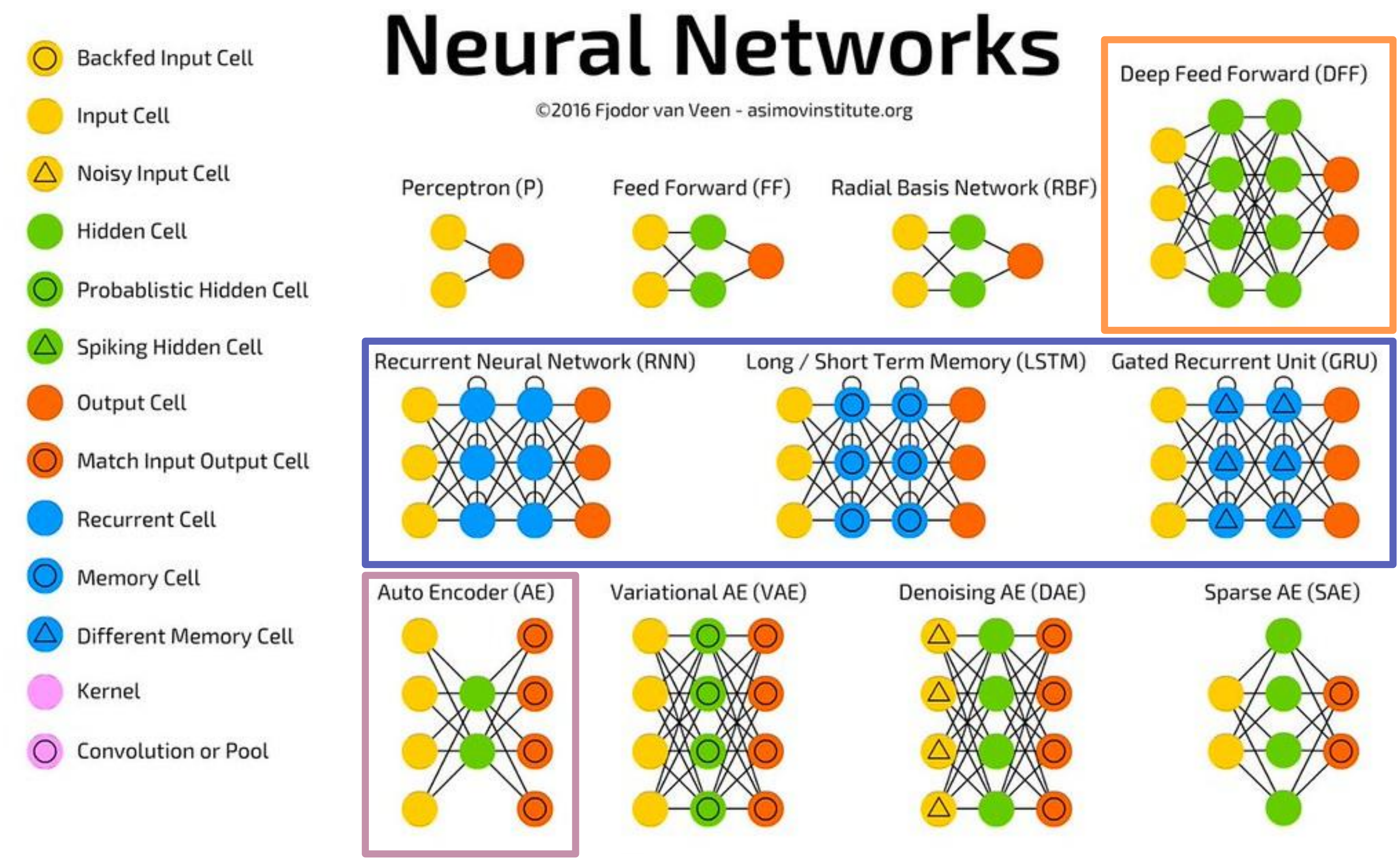
* Facilitate: Utilize other users' data till enough data about the individual is collected (cold-start problem)

Advantages of Personalized ML

- **Better performance** than generic models for the majority of individuals (Jaques et al., 2017; Suhara et al., 2017; Taylor et al., 2017; Vaizman et al., 2017; Can et al., 2019; Utsumi et al., 2019)
- Has the potential to enable **privacy-preserving**, personalized ML solutions
- Has the potential to **decrease bias** in ML models for minority populations

Novel, active research topic with growing scientific interest from major institutions (MIT Media Lab and University of Cambridge) and companies (Empatica Wearables, Affectiva)

Fundamentals: Deep Learning for Personalized ML

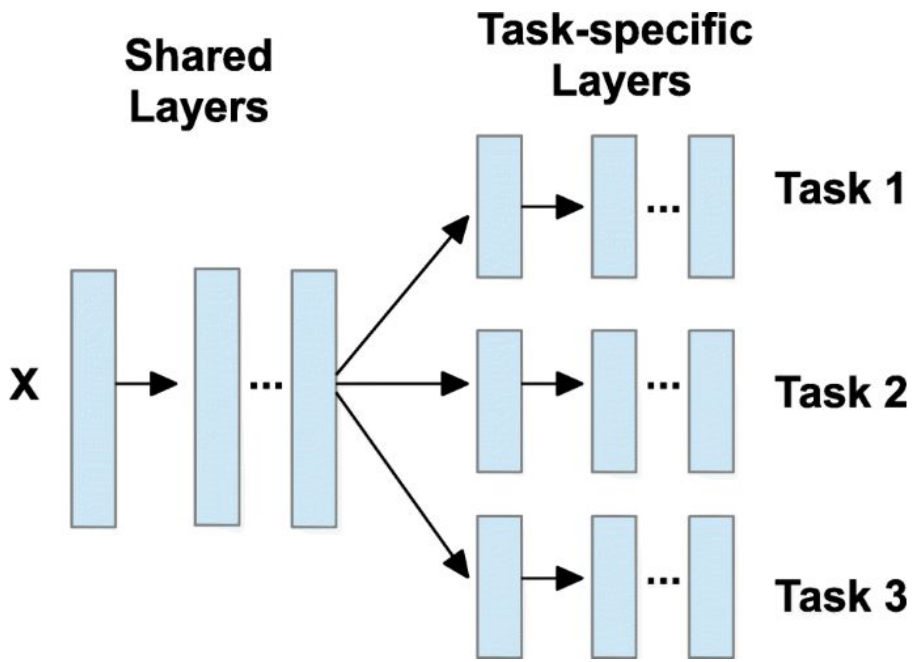


Classic Network

Sequence Modeling

Encoder - Decoder Network

Fundamentals: Multi-task Learning



Source: Thung et al. (2018)

- Subcategory of Transfer Learning
- Knowledge is shared between tasks (shared layers)
- Multiple tasks $T_1, T_2, T_3 \ldots T_n$ are trained jointly
- **Target:** Minimization of objective for all tasks

RAIS

When to use Multi-task Learning

- If you would benefit from shared low-level features
- If you have similar amounts of data for each task
- If you can train a big enough NN to do well on all tasks
- If you have absent labels

Personalized ML for Mental Wellbeing

Source: https://www.trendhunter.com/trends/sharing-app

Where it all started...

Moodscope (LiKamWa et al., 2013)

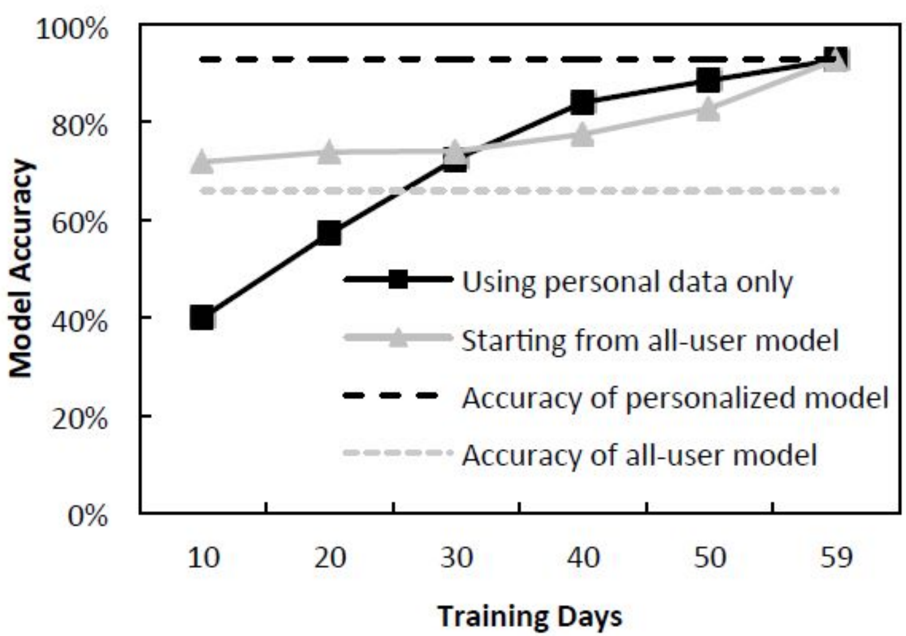


Figure 9: Pleasure training accuracy vs. training data size

Goal: Mood Inference based on Smartphone Usage Patterns

Model: Multi-linear regression (Personalized, Generic and Hybrid models)

Max Accuracy:

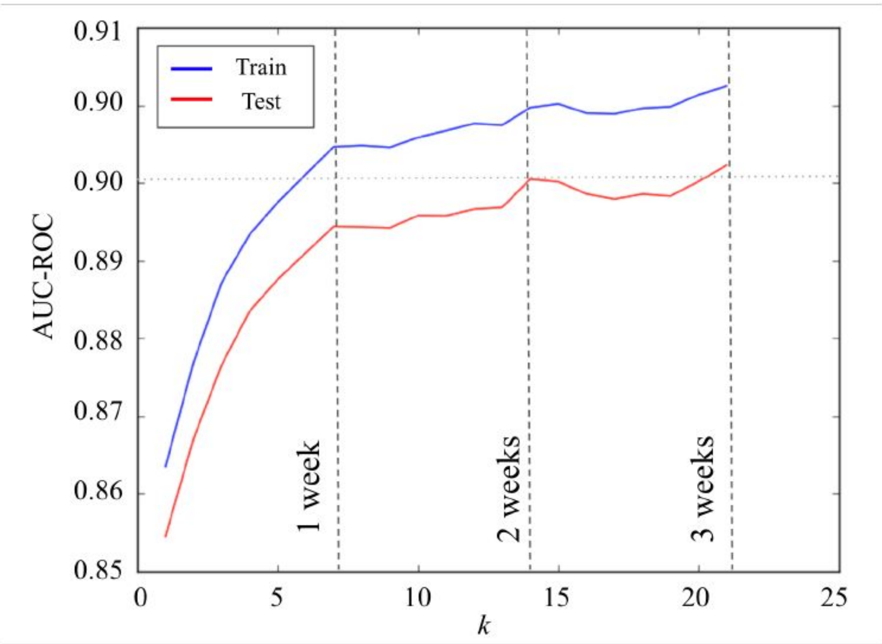
(AUC not reported)

- Generic: 66%
- Personalized: 99%

RAIS

Where Deep Learning kicked in...

DeepMood (Suhara et al., 2017)



Goal: Depression forecasting based on self-reports

Model: RNN with hidden LSTM units

AUC-ROC:

- Generic: 88.6%

What if we combined Personalized ML and Deep Learning?

Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation

Jaques et al. (2017)

Overview

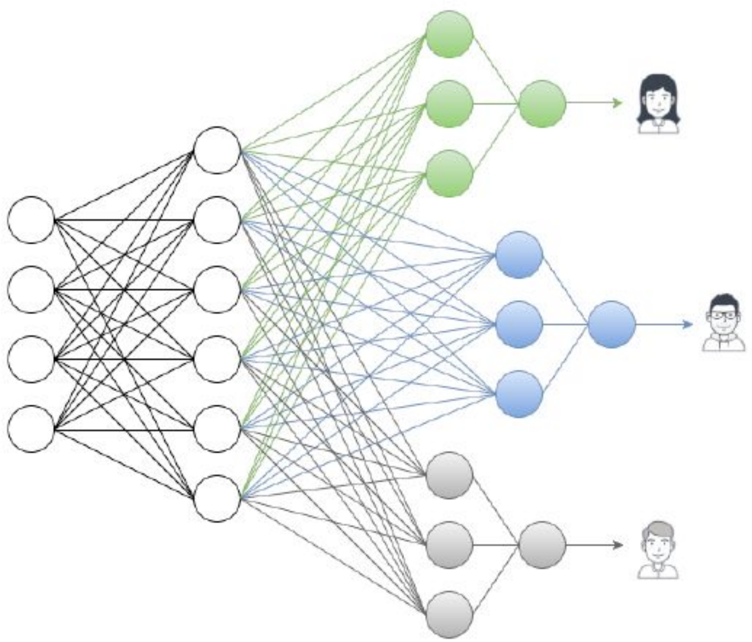
Goal:

Forecasting a person's mood from passively collected data (wearables and smartphones) and self-reported labels

Contributions:

- Taking advantage of both the data collected from the general population and the individual's data through a **multi-task, forward-feed DNN**
- **Forecasting** instead of detecting a person's mood
- Treating **mood as a regression problem** rather than a binary classification problem
- Provide considerable **performance boost** for the mood prediction problem

Model



Features: Manually designed

- Physiology (skin conductance, temperature, accelerometer; total of 342)
- Location (GPS coordinates; total of 15)
- Phone usage (SMS, calls, screen on/off; total of 75)
- Surveys (sleep, exercise, academic and extracurricular activities, etc.; total of 38)
- Weather (sunlight, temperature, barometric pressure, etc.; total of 40)
- Mood labels (mood and stress in a 1-100 range)

Task: An individual person

Training Iteration: A mini-batch consists of a single person's data and is used to predict the target labels for this person. Errors are back-propagated to update shared and task-specific layers' weights.

Results & Limitations

	Model	Mood	Stress	Health	Total
Traditional	GP	16.0	17.2	16.7	16.6
	NN	15.0	17.1	16.5	16.2
Personalized	DA-GP	14.8	16.4	**14.6**	15.3
	MTL-NN	**13.0**	**14.1**	**12.9**	**13.3**

Personalized MTL-NN provided statistically significant better performance (Mean Absolute Error - MAE) for all target labels

Limitations:

- **Cold-start problem;** no way to incorporate new users to the MTL-NN
- **Small data sample** (N=69)
- **High Label requirements**; >15 days of data required for the personalization
- **No sequence modeling**
- **Manual feature design**
- **Inability to predict** far into the future (only one step advance)

Personalized Multitask Learning for Predicting tomorrow's Mood, Health, and Stress

Taylor et al. (2017)

Overview

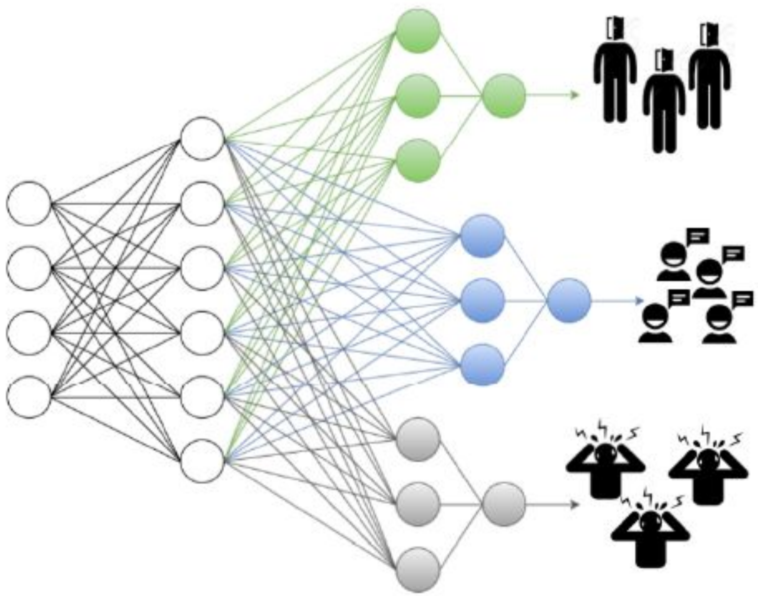
Goal:

Forecasting a person's mood from passively collected data (wearables and smartphones) and self-reported labels

Contributions:

- Taking advantage of both the data collected from the general population and the individual's data through a **multi-task, forward-feed DNN**
- **Handling cold-start problem** through user clustering (0 labels needed for new users); ability to predict future wellbeing **without requiring labels** for each person
- **Forecasting** instead of detecting a person's mood
- Provide considerable **performance boost** for the mood prediction problem

Model



Features: Manually designed

- Physiology (skin conductance, temperature, accelerometer; total of 342)
- Location (GPS coordinates; total of 15)
- Phone usage (SMS, calls, screen on/off; total of 75)
- Surveys (sleep, exercise, academic and extracurricular activities, etc.; total of 38)
- Weather (sunlight, temperature, barometric pressure, etc.; total of 40)
- Mood labels (mood and stress in a 1-100 range)

Task: A cluster of users with similar personality

Training Iteration: A mini-batch consists of a single cluster's data and is used to predict the target labels for this cluster. Errors are back-propagated to update shared and task-specific layers' weights.

RAIS

Results & Limitations

	Classifier	Mood	Stress	Health
Baseline	Majority class	50.4%, .500	50.7%, .500	54.4%, .500
	LSSVM	60.2%, .603	58.1%, .581	62.3%, .614
STL	LR	56.9%, .569	59.4%, .594	55.4%, .544
	NN	60.5%, .606	60.1%, .600	65.9%, .648
	NN (all feats)	65.8%, .658	67.9%, .678	59.0%, .591
	MTMKL	59.4%, .594	58.8%, .587	62.0%, .610
MTL - moods	HBLR	58.3%, .583	57.8%, .578	55.1%, .551
	MTL-NN	60.2%, .602	60.1%, .600	65.3%, .643
	MTL-NN (all feats)	67.0%, .670	68.2%, .682	63.0%, .623
	MTMKL	78.7%, .787	77.6%, .776	78.7%, .786
MTL - people	HBLR	72.0%, .720	73.4%, .734	76.1%, .760
	MTL-NN	77.6%, .776	78.6%, .785	79.7%, .792
	MTL-NN (all feats)	78.4%, .784	81.5%, .815	82.2%, .818

Personalized MTL-NN provided AUC-ROC of ~78% for mood prediction compared to ~65% for a generic NN

Limitations:

- **Cold-start problem**; new users need to complete a personality scale for the MTL-NN model
- Target variable is **binary**; **removal** of most **ambiguous users**
- Relatively **small data sample** (N=104)
- **No sequence modeling**
- **Manual feature design**
- **Inability to predict** far into the future (only one step advance)

Sequence Multi-task Learning to Forecast Mental Wellbeing from Sparse Self-reported Data

Spathis et al. (2019)

Overview

Goal:

Forecasting a person's future sequences of mood from passively collected data (wearables) and sparse, self-reported labels

Contributions:

- Predicts **multiple steps ahead**; not just one
- **Multi-task learning** utilized to predict different dimensions of mood
- **Sequence modeling** utilized through LSTM units
- **Automated feature extraction** through seq2seq encoder-decoder model
- **Performance boost** over single-task alternative and traditional ML approaches for the mood prediction domain

Model

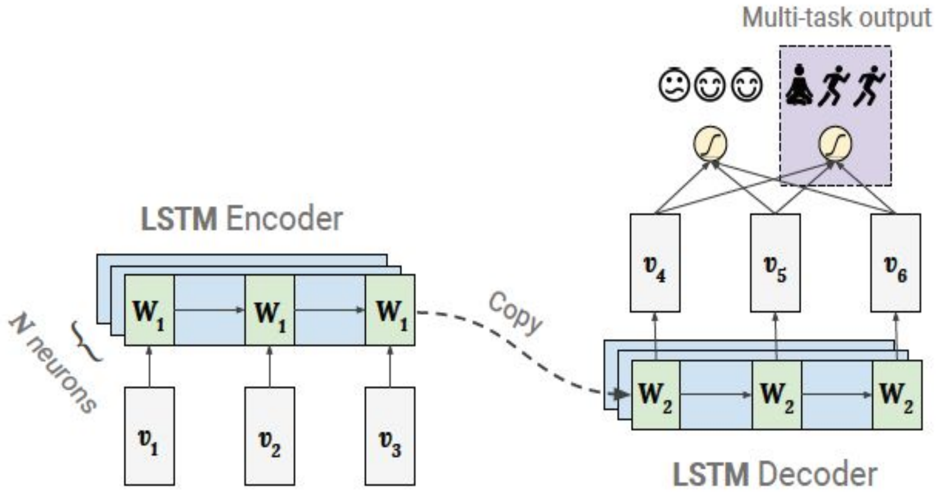


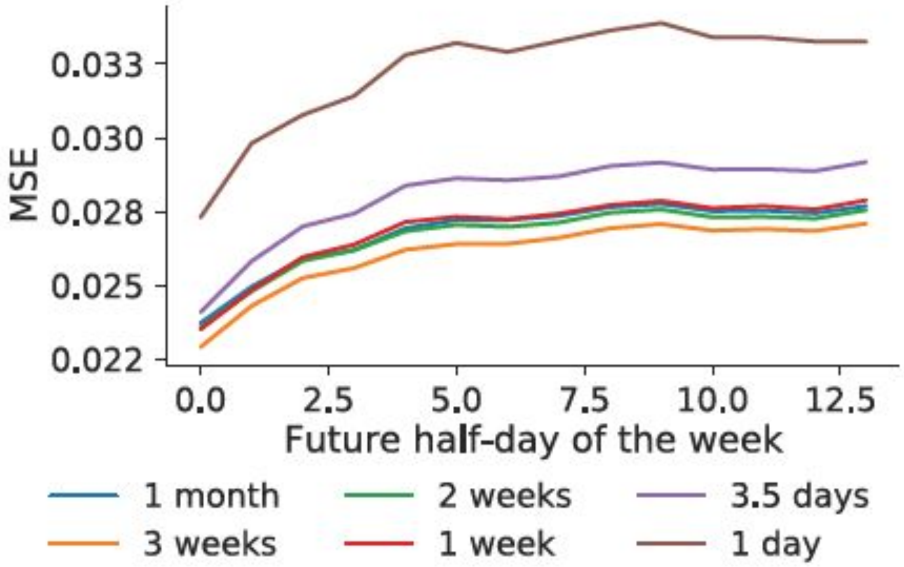
Figure 3: LSTM Encoder-Decoder model. The mood sequence (v_1,v_2,v_3) passes through an LSTM (states W_1), gets transformed to a single vector (dotted) and decoded through another LSTM (W_2) that predicts future mood sequences (v_4,v_5,v_6). Two fully-connected layers are applied to every time-step of the output (yellow circle), one for valence and one for arousal (purple box).

Task: A dimension (valence/arousal) of affect

Training Iteration: Pass the input through a standard **LSTM layer as an Encoder** in order to map the past mood into a **fixed length representation with the size of the prediction**, and then another **LSTM layer as a Decoder to reconstruct the original** sequence in future steps

RAIS

Results & Limitations



The MTL model offers **statistically significant performance boost** in predicting both valence and arousal over the naive baseline, the SVR, and the GBR ($p < 0.001$).

3 weeks of data offer the best performance; the **error increases** the **more days in the future** we are trying to **predict**

Limitations:

- **Lack of personalization:** Does not explore the concept of personalized ML; potentially could offer greater performance boost
- **Lack of baselines:** Does not offer comparisons with previous works
- **Label requirements:** Best results are achieved with 3 weeks of labeled data (sparse)

RAIS

Future Work Directions

- Personalized ML and **Sequence Modeling**
- Personalized ML and **Algorithmic Bias**
- **Privacy-preserving** Personalized ML
- Personalized ML and Multi-task Learning: The **Cold-start Problem**
- **Interpretability** of Personalized ML Models
- Personalized ML: Quantifying **Uncertainty**
- **New sub-domains** within the health and wellbeing domain
- **Concept Drift Adaptation**:
  Online learning for handling concept drift

Key takeaways

- In the domain of **health and wellbeing**, personalized ML can offer a **significant performance boost**
- **Deep Neural Networks** are the **state-of-the-art** in personalized ML, but more exploration is required in the direction of **sequence modeling and interpretability**
- **Multi-task learning** and **transfer learning** have been proposed as a **solution to the cold-start problem** for new users
- Personalized ML offers a lot of **possibilities for exploration** in terms of privacy preservation, uncertainty estimation and algorithmic bias

Bibliography

Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. Sensors, 19(8), 1849.

Jaques, N., Taylor, S., Sano, A., & Picard, R. (2017, September). Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In IJCAI 2017 Workshop on artificial intelligence in affective computing (pp. 17-33). PMLR.

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013, June). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (pp. 389-402).

Rudovic, O., Lee, J., Dai, M., Schuller, B., & Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, *3*(19).

Spathis, D., Servia-Rodriguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019, July). Sequence multi-task learning to forecast mental wellbeing from sparse self-reported data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2886-2894).

Suhara, Y., Xu, Y., & Pentland, A. S. (2017, April). Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In Proceedings of the 26th International Conference on World Wide Web (pp. 715-724).

Taylor, S., Jaques, N., Nosakhare, E., Sano, A., & Picard, R. (2017). Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Transactions on Affective Computing, 11(2), 200-213.

Thung, K. H., & Wee, C. Y. (2018). A brief review on multi-task learning. Multimedia Tools and Applications, 77(22), 29705-29725.

Utsumi, Y., Guerrero, R., Peterson, K., Rueckert, D., & Picard, R. W. (2019, October). Meta-weighted gaussian process experts for personalized forecasting of AD cognitive changes. In Machine learning for healthcare conference (pp. 181-196). PMLR.

Vaizman, Y., Ellis, K., & Lanckriet, G. (2017). Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE pervasive computing, 16(4), 62-74.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S.,, & Campbell, A. T. (2014, September). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp. 3-14).

Beneficiaries / Partners

BENEFICIARIES

PARTNERS

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Innovative Training Networks (ITN) - RAIS No 813162