ANR/DFG Project MADRAS

Multi-Agent Modelling of Dense Crowd Dynamics: Predict & Understand

Université Lyon 1
Université Toulouse 1 Capitole
Bergische Universität Wuppertal
Forschungszentrum Jülich

www.madras-crowds.eu

— WP 1 —

Predicting pedestrian trajectories using neural networks

Predicting pedestrian trajectories in dense crowds

- **Simulation tools of pedestrian dynamics:** evacuation, organisation of large events, infrastructure design (urban areas, public building, train stations), etc.
- **Dense situation** (density ≥ 2 ped/m^2): exclusion and proxemic conventions + collision/conflict avoidance behaviors including anticipation mechanisms

- **Data-based predicting approaches**
  - Machine learning algorithms, typically neural network, trained on large amount of data
  - Models for Understanding versus Models for Prediction by Gilbert Saporta (COMPSTAT 2008, pp. 315–322, see also the presentation here)
- **Motivations**
  - Complex human behaviors influenced by many factors (neighborhood, type of geometry, context, etc.)
  - Always more pedestrian trajectory data publicly available
  - Recent publications present high quality of prediction

$\Rightarrow$ Cf. MADRAS

Predicting pedestrian trajectories with data-based algorithms

- **Data-based algorithms** $f_\theta(\cdot)$, for a pedestrian n,

$$\mathbf{Y}_n = f_\theta(\mathbf{X}_n) + \varepsilon_n$$

with

$\mathbf{Y}_n$	Variable to predict (trajectory over a horizon time)
$\mathbf{X}_n$	Predicting variables (e.g. relative trajectories with the neighboring pedestrians)
θ	Coefficients (parameters) to fit
ε_n	Unpredicted part (noise with volatility $\sigma > 0$)

- **High dimensions** of $\mathbf{Y}$, $\mathbf{X}$ and θ (multi-variate statistical models)
- $f_\theta : \mathbf{X} \mapsto f_\theta(\mathbf{X})$ is in general a **non-linear function** (black box), e.g. a neural network, whose coefficients θ have no direct interpretations
- **Fitting of the coefficients θ:** Training (learning) of the algorithms on large amount of data by minimising prediction errors (i.e. noise volatility)

Succinct state-of-the-art

- **Data-based algorithms predicting pedestrian dynamics: Hot topic !**
  - Surveys : Chraibi et al. [2018], Ridel et al. [2018], Bighashdel and Dubbelman [2019], Camara et al. [2020], Rudenko et al. [2020], Li et al. [2020]
  - TrajNet++ `aicrowd.com/challenges/trajnet-a-trajectory-forecasting-challenge`
- Image/video analysis : Convolutional neural networks (e.g. pedestrian detection and intend prediction for autonomous vehicles)
- Trajectory analysis : **Interaction module + recurrent LSTM neural networks**
  - Feed-forward RNN : Das et al. [2015], Ma et al. [2016], Chen et al. [2017], Sun et al. [2020]
  - Long-Short Term Memory: Alahi et al. [2016], Bisagno et al. [2018], Xue et al. [2018], Shi et al. [2019], Zhang et al. [2019], Chen et al. [2020], Zhou et al. [2021]
  - Interaction module (social grid/graph)
    - Social pooling (grid) — Alahi et al. [2016]
    - Graph recurrent network — Haddad and Lam [2020]
    - Graph convolutional network — Chen et al. [2020], Sun et al. [2020]
    - Generative adversarial network — Gupta et al. [2018], Amirian et al. [2019]
    - ... — See the review by Kothari et al. [2021]

Succinct state-of-the-art

- **Data-based algorithms predicting pedestrian dynamics: Hot topic !**
  - Surveys : Chraibi et al. [2018], Ridel et al. [2018], Bighashdel and Dubbelman [2019], Camara et al. [2020], Rudenko et al. [2020], Li et al. [2020]
  - TrajNet++ `aicrowd.com/challenges/trajnet-a-trajectory-forecasting-challenge`
- ~~Image/video analysis : Convolutional neural networks (e.g. pedestrian detection and intend prediction for autonomous vehicles)~~
- Trajectory analysis : **Interaction module + recurrent LSTM neural networks**
  - Feed-forward RNN : Das et al. [2015], Ma et al. [2016], Chen et al. [2017], Sun et al. [2020]
  - Long-Short Term Memory: Alahi et al. [2016], Bisagno et al. [2018], Xue et al. [2018], Shi et al. [2019], Zhang et al. [2019], Chen et al. [2020], Zhou et al. [2021]
  - Interaction module (social grid/graph)
    - Social pooling (grid) Alahi et al. [2016]
    - Graph recurrent network Haddad and Lam [2020]
    - Graph convolutional network Chen et al. [2020], Sun et al. [2020]
    - Generative adversarial network Gupta et al. [2018], Amirian et al. [2019]
    - ... See the review by Kothari et al. [2021]

Social grid pooling (Alahi et al. [2016])[1]

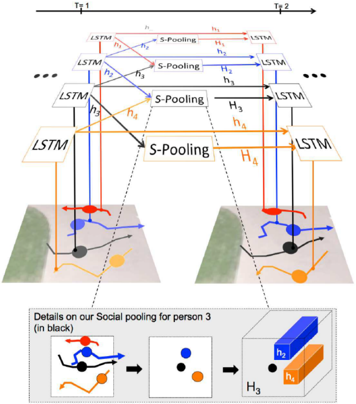


Figure 2. Overview of our Social-LSTM method. We use a separate LSTM network for each trajectory in a scene. The LSTMs are then connected to each other through a Social pooling (S-pooling) layer. Unlike the traditional LSTM, this pooling layer allows spatially proximal LSTMs to share information with each other. The variables in the figure are explained in Eq. 2. The bottom row shows the S-pooling for one person in the scene. The hidden-states of all LSTMs within a certain radius are pooled together and used as an input at the next time-step.

[1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In IEEE ICCV Conference, pages 961–971, 2016

Benchmark analysis by Alahi et al. [2016][2]

Metric	Methods	Lin	LTA	SF [73]	IGP* [60]	LSTM	our O-LSTM	our Social-LSTM
Avg. disp. error	ETH [49]	0.80	0.54	0.41	**0.20**	0.60	0.49	0.50
	HOTEL [49]	0.39	0.38	0.25	0.24	0.15	**0.09**	0.11
	ZARA 1 [39]	0.47	0.37	0.40	0.39	0.43	**0.22**	**0.22**
	ZARA 2 [39]	0.45	0.40	0.40	0.41	0.51	0.28	**0.25**
	UCY [39]	0.57	0.51	0.48	0.61	0.52	0.35	**0.27**
	Average	0.53	0.44	0.39	0.37	0.44	0.28	**0.27**
Avg. non-linear disp. error	ETH [49]	0.95	0.70	0.49	0.39	0.28	**0.24**	0.25
	HOTEL [49]	0.55	0.49	0.38	0.34	0.09	**0.06**	0.07
	ZARA 1 [39]	0.56	0.39	0.41	0.54	0.24	**0.13**	**0.13**
	ZARA 2 [39]	0.44	0.41	0.39	0.43	0.30	0.20	**0.16**
	UCY [39]	0.62	0.57	0.54	0.62	0.31	0.20	**0.16**
	Average	0.62	0.51	0.44	0.46	0.24	0.17	**0.15**
Final disp. error	ETH [49]	1.31	0.77	0.59	**0.43**	1.31	1.06	1.07
	HOTEL [49]	0.55	0.64	0.37	0.37	0.33	**0.20**	0.23
	ZARA 1 [39]	0.89	0.66	0.60	0.39	0.93	**0.46**	0.48
	ZARA 2 [39]	0.91	0.72	0.68	0.42	1.09	0.58	**0.50**
	UCY [39]	1.14	0.95	0.78	1.82	1.25	0.90	**0.77**
	Average	0.97	0.74	**0.60**	0.69	0.98	0.64	0.61

Table 1. Quantitative results of all the methods on all the datasets. We present the performance metrics as follows: First 6 rows are the Average displacement error, row 7 to 12 are the Average displacement error for non-linear regions, and the final 6 rows are the Final displacement error. All methods forecast trajectories for a fixed period of 4.8 seconds. (*) Note that IGP uses the intended ground truth destination of a person during test time unlike other methods.

Lin: Linear model (linear acceleration); LTA: Collision avoidance; SF: Social force (with group affinity and predicted destinations);
IGP: Iterative Gaussian Process (information about the final destination); LSTM: Vanilla LSTM (no interaction);
O-LSTM (interaction with direct neighbors). Social-LSTM: trained interaction weight (hidden social state).

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In IEEE ICCV Conference, pages 961–971, 2016

Example: Socially-Aware GAN (Gupta et al. [2018])[3]

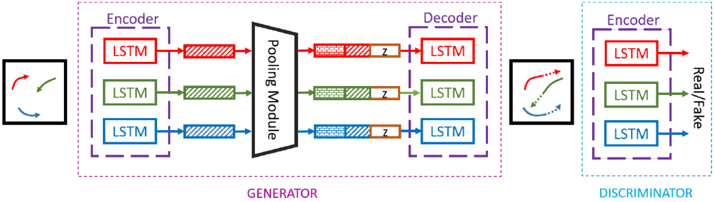


Figure 2: System overview. Our model consists of three key components: Generator (G), Pooling Module, and Discriminator (D). G takes as input past trajectories X_i and encodes the history of the person i as H_i^t. The pooling module takes as input all $H_i^{t_{obs}}$ and outputs a pooled vector P_i for each person. The decoder generates the future trajectory conditioned on $H_i^{t_{obs}}$ and P_i. D takes as input T_{real} or T_{fake} and classifies them as socially acceptable or not (see Figure 3 for PM).

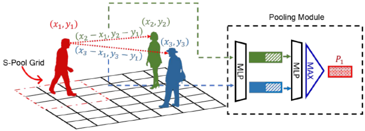
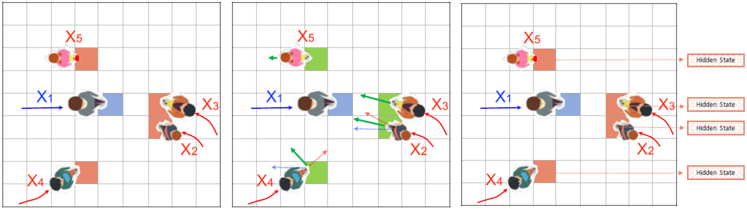


Figure 3: Comparison between our pooling mechanism (red dotted arrows) and Social Pooling [1] (red dashed grid) for the red person. Our method computes relative positions between the red and all other people; these positions are concatenated with each person's hidden state, processed independently by an MLP, then pooled elementwise to compute red person's pooling vector P_1. Social pooling only considers people inside the grid, and cannot model interactions between all pairs of people.

[3] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, A. Alahi. Social GAN: Socially acceptable trajectories with generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018*

Review by Kothari et al. [2021][4]



(a) Occupancy pooling [15] (b) Directional pooling [Ours] (c) Social pooling [15]

Fig. 3: Illustration of the grid-based interaction encoding modules. (a) Occupancy pooling: each cell indicates the presence of a neighbour (b) Our proposed directional pooling: each cell contains the relative velocity of the neighbour with respect to the primary pedestrian. (c) Social pooling: each cell contains the LSTM hidden-state of the neighbour. The constructed grid tensors are passed through an MLP-based neural network to obtain the interaction vector.

Acronym (P-Q-R-S)	Input (P)	Embed-I (Q)	Aggreg. (R)	Embed-II (S)	References
O-Grid	Position	None	Grid	MLP	**O-LSTM** [15], [45], [49]
S-Grid	H-State	None	Grid	MLP	**S-LSTM** [15], [45], [47], [51], [43], [48], [44], [59]
D-Grid	Velocity	None	Grid	MLP	**Directional Pooling [Ours]**
D-MLP-Attn-MLP	Velocity	MLP	Attn	MLP	[50]
S-MLP-Attn-MLP	H-State	MLP	Attn	MLP	**S-BiGAT** [67], [68], [64], [66], [53], [63], [65], [71], [75]
S-MLP-MaxP-MLP	H-State	MLP	MaxPool	MLP	**S-GAN** [52]
D-MLP-ConC-MLP	Velocity	MLP	Concat	MLP	[57], [58]
D-MLP-SumP-LSTM	Velocity	MLP	SumPool	MLP	**Trajectron** [55] [1]
O-LSTM-Att-MLP	Position	LSTM	Attn	MLP	**S-Attn** [82], [70]
D-MLP-ConC-LSTM	Velocity	MLP	Concat	LSTM	**DirectConcat [Ours]**

P-Q-R-S acronym for the various designs of interaction modules: P denotes the input to the interaction module, Q denotes the state embedding module, R denotes the information aggregation mechanism and S denotes aggregated vector embedding module.

[4] P. Kothari, S. Kreiss, and A. Alahi. Human trajectory forecasting in crowds: A deep learning perspective. arXiv preprint arXiv:2007.03639, 2021

Benchmark analysis (Kothari et al. [2021])[5]

Model (Acronym)	ADE/FDE	Col-I	Col-II
Grid based methods			
Vanilla	0.32/0.62	19.0	7.1
O-Grid [15]	0.27/0.52	11.7	4.9
S-Grid [15]	0.24/0.50	**2.2**	**4.6**
D-Grid [**Ours**]	0.24/0.49	**2.2**	4.8
Non-Grid based methods			
S-MLP-MaxP-MLP [52]	0.27/0.52	6.4	**5.2**
S-MLP-Attn-MLP [67]	0.26/0.52	3.7	5.4
D-MLP-SumP-LSTM [55]	0.29/0.57	13.8	6.6
O-LSTM-Attn-MLP [82]	0.24/0.48	0.8	5.2
D-MLP-MaxP-MLP	0.28/0.55	14.3	6.1
D-MLP-Attn-MLP	0.27/0.52	8.1	**5.0**
D-MLP-ConC-MLP	0.25/0.50	1.3	5.6
D-MLP-ConC-LSTM [**Ours**]	0.24/0.48	**0.6**	5.3

TABLE II: Unimodal Comparison of interaction encoder designs when forecasting 12 future time-steps, given the previous 9 time-steps, on TrajNet++ synthetic dataset. Errors reported are ADE / FDE in meters, Col I / Col II reported in %. We emphasize that our goal is to reduce Col-I without compromising distance-based metrics.

[5] P. Kothari, S. Kreiss, and A. Alahi. Human trajectory forecasting in crowds: A deep learning perspective. arXiv preprint arXiv:2007.03639, 2021

Pedestrian trajectory data

- **Naturalistic publicly available trajectory data-sets**

  https://github.com/crowdbotp/OpenTraj   Amirian et al. [2020]

  - Train station, urban areas, university surrounds (!)
  - Low or medium density level (lower than 1 or 2 ped/m^2)
  - Most of literature trained on naturalistic trajectory data-sets

- **Experimental trajectory data-sets**

  https://ped.fz-juelich.de/database

  - Reproducible experiments done in laboratory conditions, clear context and pedestrian intend, high resolution, no missing data
  - High density level (cf. Basigo project with density up to 8 ped/m^2)

- **Synthetic trajectory data-sets**

  - Trajectory data obtained by simulation (multi-agent models, social force model, etc.)
  - Hybrid approach

Methodology

- **Benchmark analysis**
  - Multiple linear regression — linear reference
  - Recurrent neural networks — non-linear reference

    + LSTM term + Interaction module + Types of inputs

    - Interaction: Selection/weighting of the neighboring pedestrians/obstacles (social pooling)
    - Type of inputs: Past relative trajectories over given time interval, spacing variables (distance, velocity difference, time-to-collision, bearing angle, etc.)
    - Override the social pooling using pertinent variables and anticipation mechanisms (e.g. TTC)?

- **Bootstrap cross-validation**
  - Sub-sampling the data in training and testing sets in bootstrap loops
    1. Setting the coefficients by iterative error minimisation (back-propagation) on training sets
    2. Compute the error on testing sets
  - Setting the optimal algorithmic complexity by minimising the mean testing error
  - Evaluation of the estimation precision of the error using bootstrap interval

Challenges

- **Setting of the algorithm architecture/complexity and interaction graph**
  - Deep-learning : Number of layers, inputs
  - Feed-forward (MLP) `VERSUS` LSTM
  - Social pooling `VERSUS` Distance variables and anticipation mechanisms

- **Robustness and reliability of the predictions**
  - Testing for high density levels
  - Testing for new data, scenes, different geometries, density levels
  - Analysis of prediction horizon time

- **Hybrid approach**
  - Training the networks on simulated data of simple physic-based models
  - Ability of the networks to capture physical rules
  - "Network complexity" necessary to capture physical rules

Possible time-schedule

	2021			2022				2023				
	2	3	4	1	2	3	4	1	2	3	4	1
`WP 1.1`												
Literature review	■	■	■	■	■	■						
Real/experimental/simulated data		■	■	■	■							
Preliminary data analysis			■	■	■	■						
`WP 1.2`												
MLP vs. LSTM			■	■	■	■	■	■	■	■		
Social pooling vs. Anticipation			■	■	■	■	■	■	■	■		
Testing physical models			■	■	■	■	■	■	■	■		
`WP 1.3`												
Comparison with physical models					■	■	■	■	■	■	■	■
Model validation							■	■	■	■	■	■
Implementation as MAS in GAMA							■	■	■	■	■	■
Fête des Lumières							■	■	■	■	■	■

References I

A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In IEEE ICCV Conference, pages 961–971, 2016.

J. Amirian, J.-B. Hayet, and J. Pettre. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

J. Amirian, B. Zhang, F. V. Castro, J. J. Baldelomar, J.-B. Hayet, and J. Pettre. Opentraj: Assessing prediction complexity in human trajectories datasets. In Proceedings of the Asian Conference on Computer Vision (ACCV), 2020.

A. Bighashdel and G. Dubbelman. A survey on path prediction techniques for vulnerable road users: From traditional to deep-learning approaches. In IEEE Intelligent Transportation Systems Conference (ITSC), pages 1039–1046, 2019.

N. Bisagno, B. Zhang, and N. Conci. Group LSTM: Group trajectory prediction in crowded scenarios. In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

F. Camara, N. Bellotto, S. Cosar, et al. Pedestrian models for autonomous driving part II: high-level models of human behavior. IEEE Transactions on Intelligent Transportation Systems, 2020.

K. Chen, X. Song, D. Han, J. Sun, Y. Cui, and X. Ren. Pedestrian behavior prediction model with a convolutional LSTM encoder–decoder. Physica A: Statistical Mechanics and its Applications, 560:125132, 2020.

References II

Y. Chen, M. Everett, M. Liu, and J. P. How. Socially aware motion planning with deep reinforcement learning. In IEEE IROS Conference, pages 1343–1350, 2017. .

M. Chraibi, A. Tordeux, A. Schadschneider, and A. Seyfried. Pedestrian and Evacuation Dynamics - Modelling (2nd Edition). Springer, Berlin, 2018. .

P. Das, M. Parida, and V. K. Katiyar. Analysis of interrelationship between pedestrian flow parameters using artificial neural network. J. Mod. Transport., 23(4):298–309, 2015. .

A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social GAN: Socially acceptable trajectories with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. .

S. Haddad and S.-K. Lam. Self-growing spatial graph network for context-aware pedestrian trajectory prediction. arXiv preprint arXiv:2012.06320, 2020. .

P. Kothari, S. Kreiss, and A. Alahi. Human trajectory forecasting in crowds: A deep learning perspective. arXiv preprint arXiv:2007.03639, 2021. .

J. Li, H. Ma, Z. Zhang, and M. Tomizuka. Social-WAGDAT: Interaction-aware trajectory prediction via wasserstein graph double-attention network. arXiv preprint arXiv:2002.06241, 2020. .

Y. Ma, E. W. M. Lee, and R. K. K. Yuen. An artificial intelligence-based approach for simulating pedestrian movement. IEEE Trans. Intell. Transp. Syst., 17(11):3159–3170, 2016. .

D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf. A literature review on the prediction of pedestrian behavior in urban scenarios. In IEEE International Conference on Intelligent Transportation Systems (ITSC), pages 3105–3112, 2018. .

References III

A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. Human motion trajectory prediction: a survey. The International Journal of Robotics Research, 39(8): 895–935, 2020. ↗.

X. Shi, X. Shao, Z. Guo, G. Wu, H. Zhang, and R. Shibasaki. Pedestrian trajectory prediction in extremely crowded scenarios. Sensors, 19(5):1223, 2019. ↗.

J. Sun, Q. Jiang, and C. Lu. Recursive social behavior graph for trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 660–669, 2020. ↗.

H. Xue, D. Q. Huynh, and M. Reynolds. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. In IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1186–1194, 2018. ↗.

P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng. SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. ↗.

Y. Zhou, H. Wu, H. Cheng, K. Qi, K. Hu, C. Kang, and J. Zheng. Social graph convolutional LSTM for pedestrian trajectory prediction. IET Intelligent Transport Systems, 2021. ↗.