DAG-Net: Double Attentive Graph Neural Network for Trajectory Forecasting

A. Monti, A. Bertugli, S. Calderara, R. Cucchiara

AImageLab - University of Modena and Reggio Emilia, Modena, Italy | alessio.monti@unimore.it

ICPR 2020 — 25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION — Milan, Italy 10 | 15 January 2021

1. Motivation

Trajectory forecasting has become an important component for several downstream applications: intelligent transportation could leverage insights about human motion to preventively avoid dangerous situations, while surveillance systems could use this information to improve crowd control.
The task finds several demands also outside the smart cities context, for example in team sports, where predicting the movements of the opposing team could represent a crucial factor for tactical analysis.

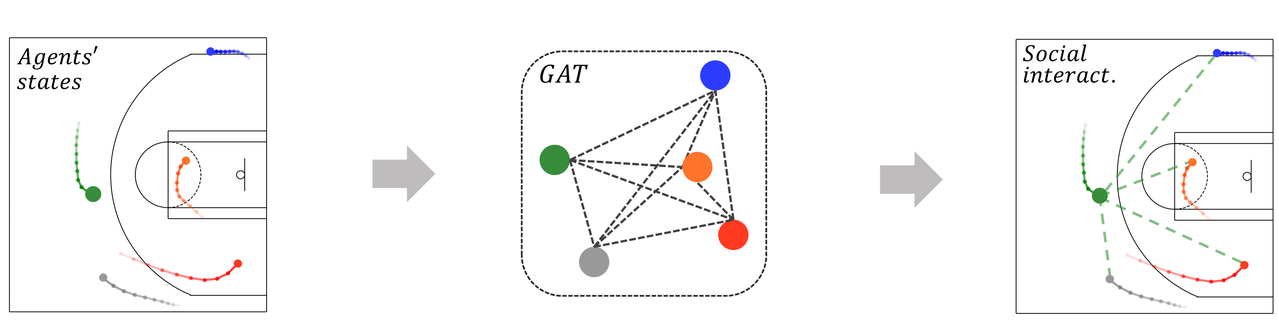
The task results particularly tough because of two peculiar characteristics of human movement:

- **multi-modality**: given a brief history of past positions, an agent could move in several plausible (and equally correct!) ways
- **social interactions**: people plan their path reading each other future possible behaviours, thus being heavily influenced by the surrounding subjects

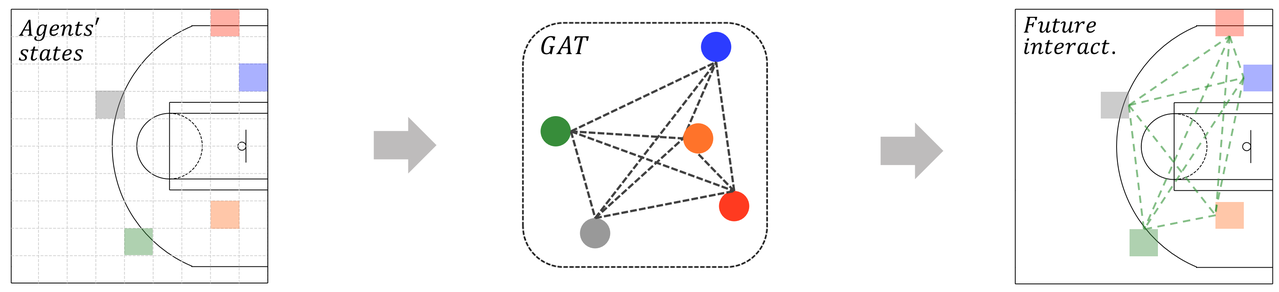
2. DAG-Net

Our baseline is represented by the **Variational Recurrent Neural Network** [2], *i.e.* the combination between a Recurrent Neural Network and a generative model like the Variational Auto-Encoder. By conditioning the VAE generations on a recurrent hidden state, the model acquires impressive sequence modelling capabilities. However, by processing each sequence independently, the model is unable to leverage useful insights as *social interactions*.

To recover spatial relationships between agents that share the same scene, we firstly propose to enhance the baseline model with a **Graph Attention Network** [4] which refines the underlying recurrent hidden state with community information. This way, the model is conditioned on neighbourhood **past information** and is able to produce more likely results.



To produce even more reliable predictions, we then propose to jointly explore **future information**: we express the likely behaviour of each agent in terms of **future goals**, *i.e.* by dividing the pedestrian space in a grid of cells and by extracting the regions of space an agent will occupy in the next time-steps. We then employ a second Graph Attention Network [4] to share agents' goals across the scene neighbourhood.



Exploiting a double graph allows us to jointly explore and condition the model on both past and future information, increasing the model capabilities and producing more accurate predictions.

3. Ablation

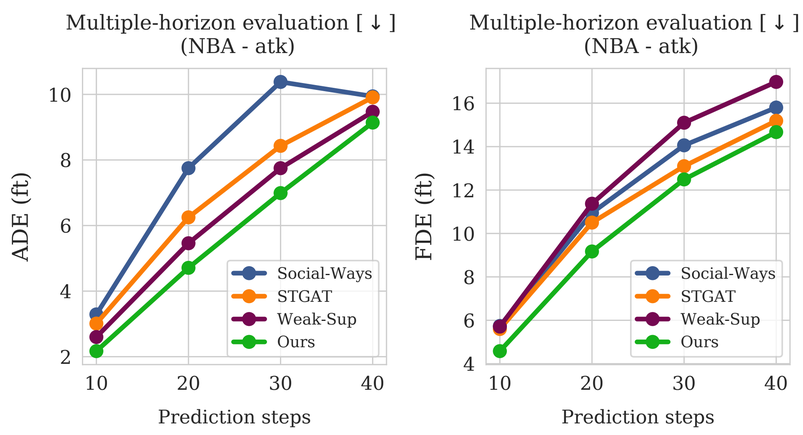
Dataset	Model	Inter.	Goals	ADE	FDE
NBA (atk)	VRNN	✗	✗	9.58	15.83
	A-VRNN	✓	✗	9.67	15.96
	DAG-Net	✓	✓	**9.18**	**13.54**
NBA (def)	VRNN	✗	✗	7.07	10.62
	A-VRNN	✓	✗	7.01	10.42
	DAG-Net	✓	✓	**7.01**	**9.76**
SDD	VRNN	✗	✗	0.58	1.17
	A-VRNN	✓	✗	0.56	1.14
	DAG-Net	✓	✓	**0.53**	**1.04**

Using a stand-alone network without further solutions does not allow to thoroughly capture the nature of human paths because of the lack of important information like social relationships.

Our Attentive-VRNN achieves better performance: still, the model is not able to capture *future* relationships between agents.

The performance achieved by DAG-Net highlight the importance of combining future objectives in a structured way and of embedding this information into the generation process.

4. Long-term comparison



To show the superior predictive capabilities granted by our solution, we test the architecture on different prediction lengths by producing **multiple-horizon evaluations**.

To this end, we evaluate our proposal against several competitors on increasing prediction splits, from 10 to 40 time-steps. DAG-Net globally outperforms the competitors in all the different evaluations, proving its strength in all the possible predictive settings.

5. Results

Model	NBA (atk) ADE	NBA (atk) FDE	NBA (def) ADE	NBA (def) FDE	SDD ADE	SDD FDE
STGAT [3]	9.94	15.80	7.26	11.28	0.58	1.11
Social-Ways [1]	9.91	15.19	7.31	10.21	0.62	1.16
Weak-Sup. [5]	9.47	16.98	7.05	10.56	-	-
Our	**8.98**	**14.08**	**6.87**	**9.76**	**0.53**	**1.04**

DAG-Net achieves impressive results also when compared to the state-of-the-art methods with the standard evaluation protocol. STGAT shows strong performances but totally lacks additional information about what could happen in the long-term.
With its attention-based pooling, Social-Ways returns similar outcomes: the combination of its geometric social features gives useful clues.
Even though its different approach, Weak Supervision gains promising results too: the exploitation of future intentions allow agents to correctly plan their future paths.
However, by jointly considering past and future information, our solution is more capable of capturing the real nature of human paths and of reaching smaller errors in particularly different domains as the urban and the sports settings.

6. References

[1] Amirian et al. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs. In *CVPRW*, 2019.

[2] Chung et al. A Recurrent Latent Variable Model for Sequential Data. In *NIPS*, 2015.

[3] Huang et al. STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction. In *ICCV*, 2019.

[4] Veličković et al. Graph Attention Networks. In *ICLR*, 2018.

[5] Zhan et al. Generating Multi-Agent Trajectories using Programmatic Weak Supervision. In *ICLR*, 2019.

Code

Project code available on Github: