Transformer Networks for Trajectory Forecasting

Francesco Giuliari[1], Irtiza Hasan[2], Prof. Marco Cristani[1], and Prof. Fabio Galasso[3]

[1] University of Verona,
[2] Inception Institute of Artificial Intelligence (IIAI) ,
[3] Sapienza University of Rome

Trajectory forecasting

Goal:

- Predict people movements

Utility:

- Person detection and tracking
- Autonomous agents

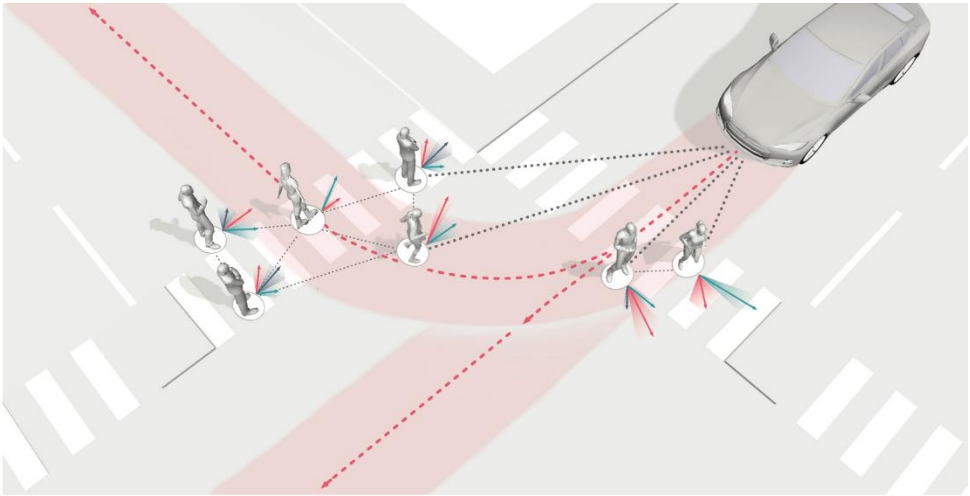
Standard solution:

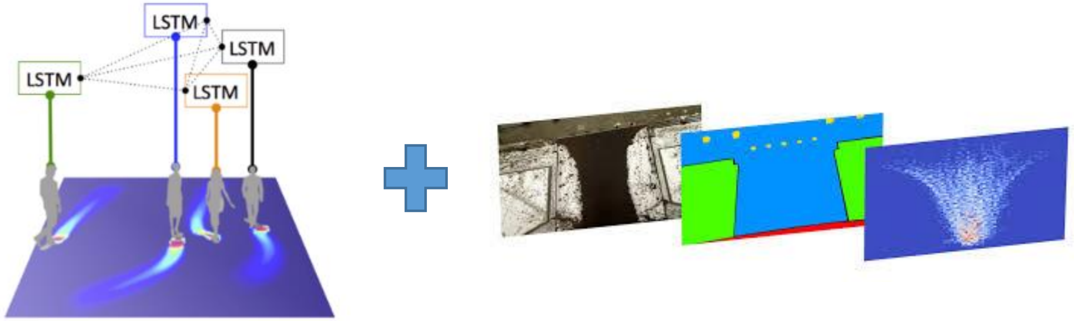
- Recurrent Neural Networks

Our proposal:

- Use the Transformer Network[1]

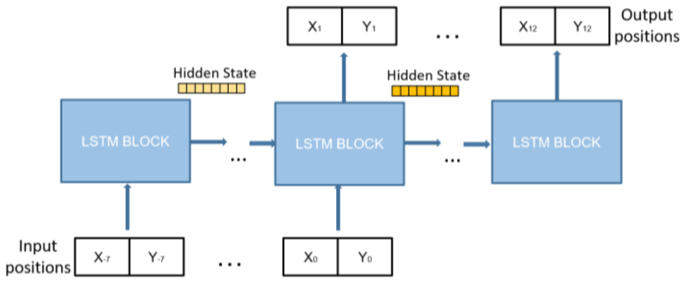
[1] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*. 2017.





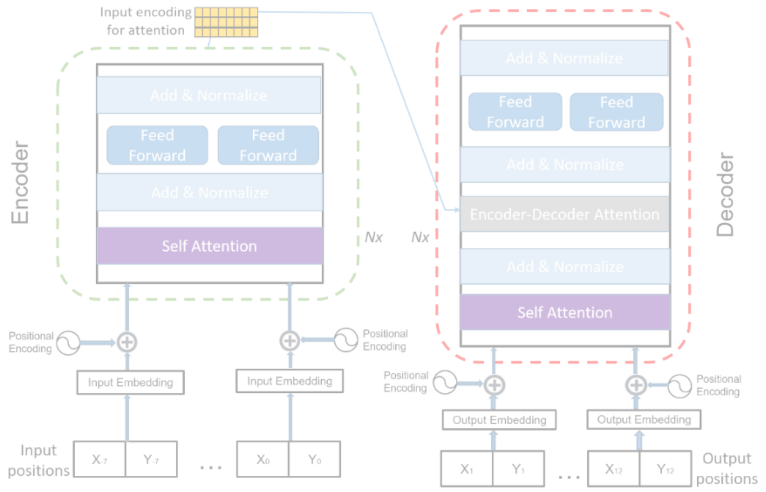
Images from ***social lstm*** *and "****Back to the Future****" blog post by Boris Ivanovic*

LSTM



- Use recurrent layers to deal with sequential data
- History encoded in a Hidden State
  - Limited memory
  - Risk of forgetting initial observations
- Observed values and previous predictions are treated in the same way
  - Amplification of prediction errors

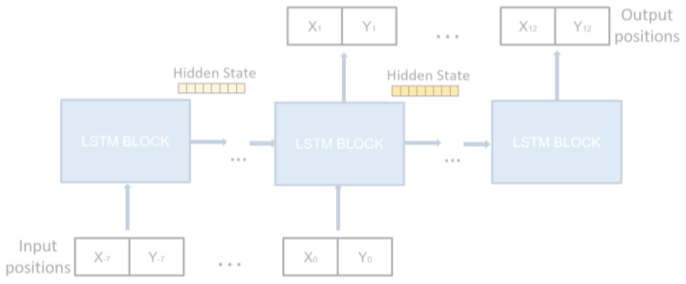
The Transformer



Transformer Network (TF)

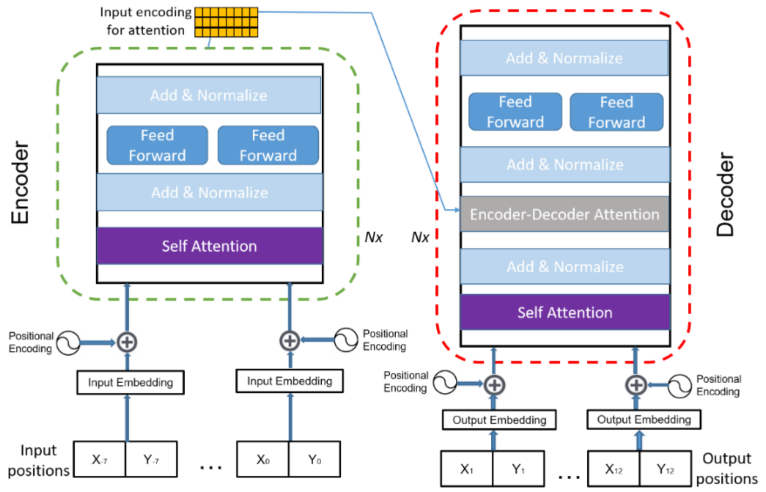
- Use attention mechanisms to deal with sequential data
- All the input elements are available during inference
  - No loss of information
- Different treatment to observed values and previous predictions
  - Prediction errors can be controlled

LSTM



- Use recurrent layers to deal with sequential data
- History encoded in a Hidden State
  - Limited memory
  - Risk of forgetting initial observations
- Observed values and previous predictions are treated in the same way
  - Amplification of prediction errors

The Transformer



Transformer Network (TF)

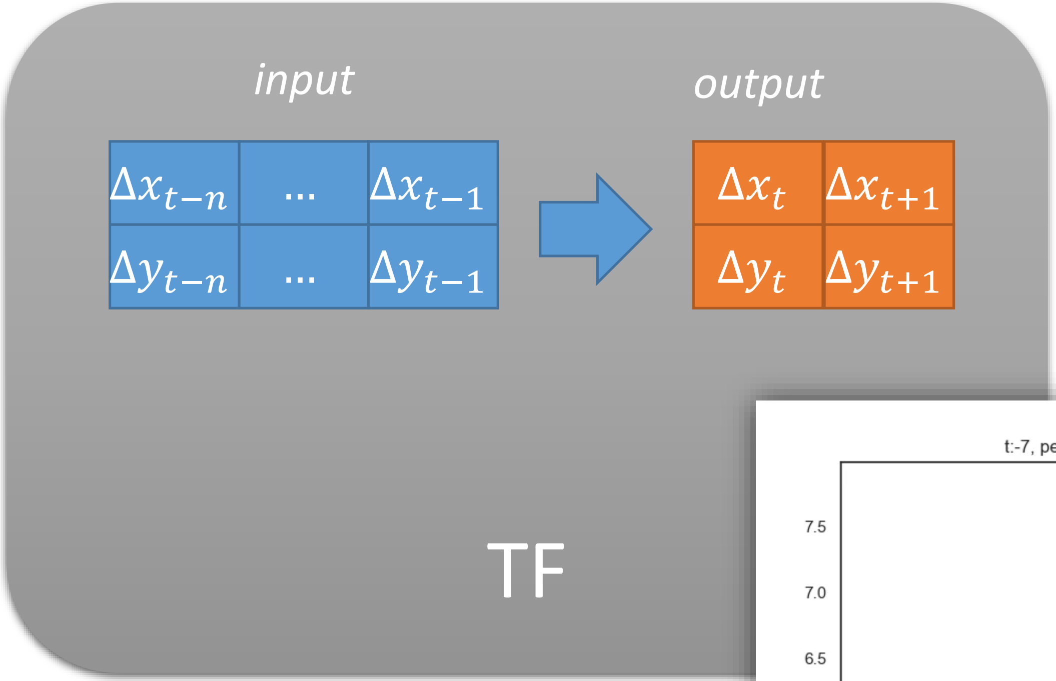
- Use attention mechanisms to deal with sequential data
- All the input elements are available during inference
  - No loss of information
- Different treatment to observed values and previous predictions
  - Prediction errors can be controlled

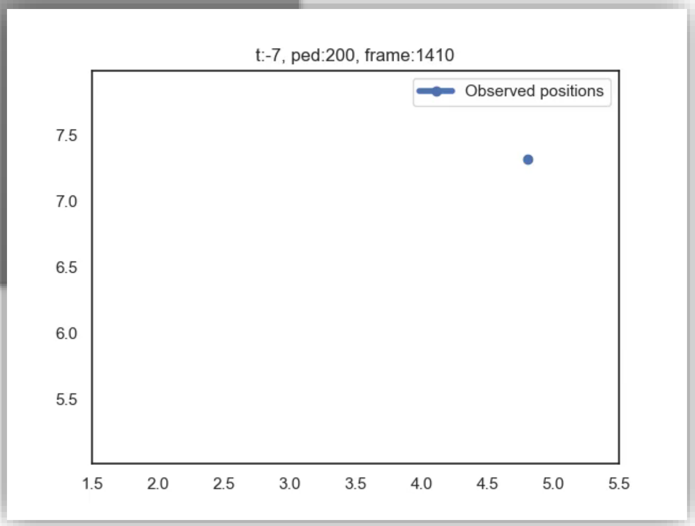
Our proposed models (1/2)

TF

- Single prediction

- Input:
  - Velocity vectors
- Output:
  - Velocity vectors





Our proposed models (2/2)

TFq

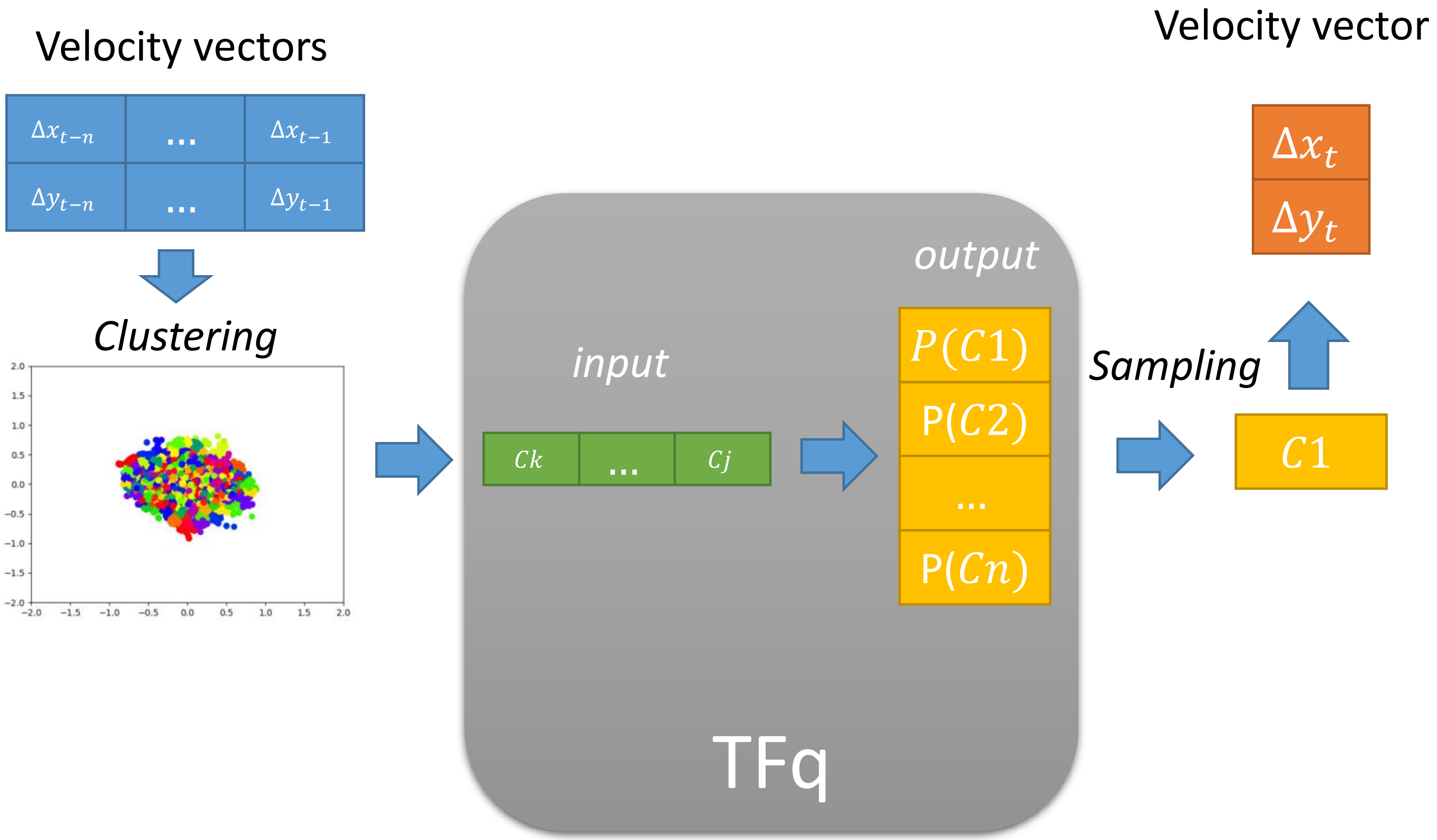
- Quantized version
- Multiple prediction

Input:

- Sequence of clusters

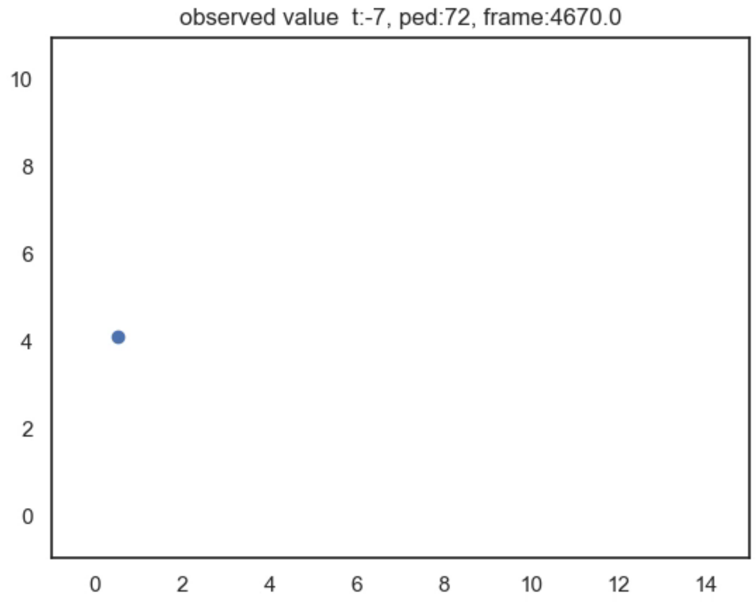
Output:

- Cluster probabilities

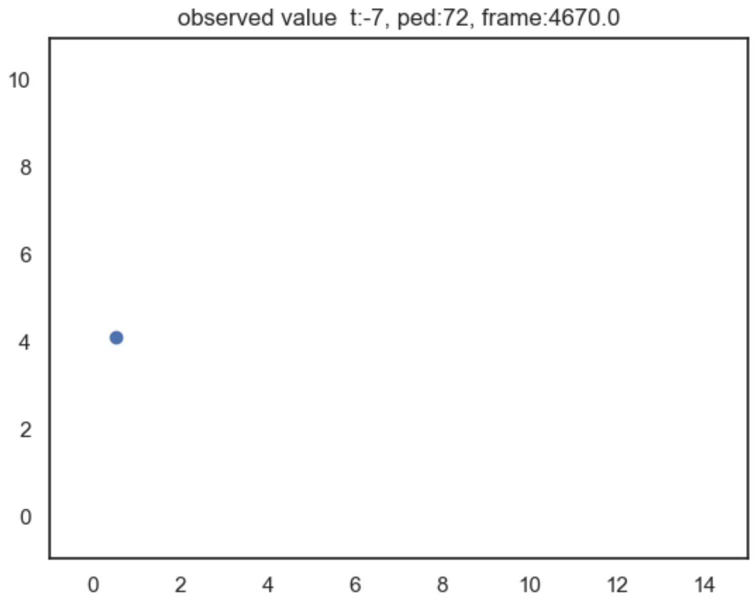


Multiple predictions

100 sampled trajectories



Kernel Density Estimation



Quantitative Results

Trajnet Benchmark

- Single prediction
- Server-side Testing
- Data from many datasets:
  - ETH
  - UCY
  - Stanford Drone Dataset
  - PETS 2009

Method	FAD	MAD	Needs social cues
TF	***1.197***	***0.356***	***no***
REDv3	1.201	0.360	no
SR-LSTM	1.261	0.37	yes
S.Forces (ewap)	1.266	0.371	yes
TF_q	*1.300*	*0.416*	*no*
BERT	*1.354*	*0.440*	*no*
BERT_NLP_pt.	*1.357*	*0.447*	*no*
MX-LSTM	1.374	0.399	yes
S.Forces (attr)	1.395	0.412	yes
LSTM	1.793	0.491	no
S-GAN	2.107	0.561	yes

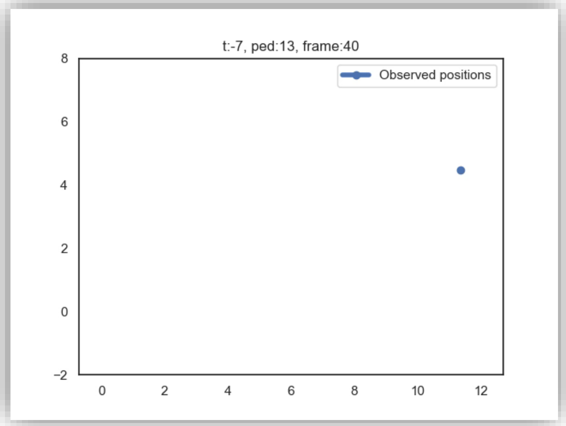
Blue italic indicates approaches proposed in this work.

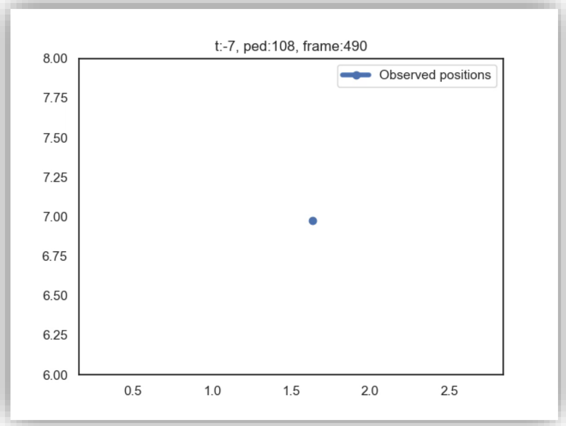
Comparison against SOTA

ETH+UCY: Single Prediction

	Linear	LSTM-based					TF-based
	Individual	Individual		Social			Individual
	Interpolat.	LSTM [10]	S-GAN-ind [10]	Social LSTM [10]	Soc. Att. [14]	Trajectron++ [22]	TF (ours)
ETH	1.33/2.94	1.09/2.94	1.13/2.21	*1.09/2.35*	*0.39/3.74*	*0.71/1.68*	**1.03/2.10**
Hotel	0.39/0.72	0.86/1.91	1.01/2.18	*0.79/1.76*	*0.29/2.64*	*0.22/0.46*	**0.36/0.71**
UCY	0.82/1.59	0.61/1.31	0.60/**1.28**	*0.67/1.40*	*0.20/0.52*	*0.41/1.07*	**0.53**/1.32
Zara1	0.62/1.21	**0.41/0.88**	0.42/0.91	*0.47/1.00*	*0.30/2.13*	*0.30/0.77*	0.44/1.00
Zara2	0.77/1.48	0.52/1.11	0.52/1.11	*0.56/1.17*	*0.33/3.92*	*0.23/0.59*	**0.34/0.76**
avg	0.79/1.59	0.70/1.52	0.74/1.54	*0.72/1.54*	*0.30/2.59*	*0.37/0.95*	**0.54/1.17**

- **TF** outperform all the methods that use the same amount of information
- **TF** outperform 2 of the 3 methods that use social cues while having access to less information





Comparison against SOTA

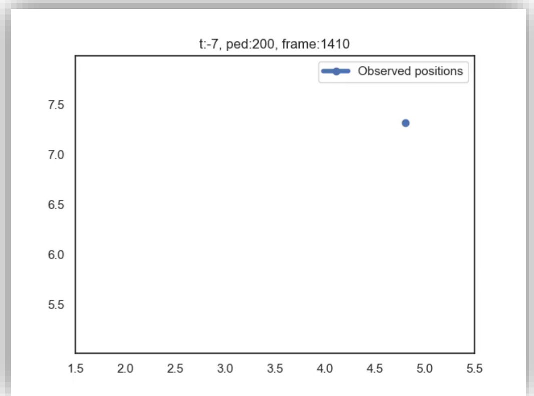
ETH+UCY: Multiple Prediction

- **TFq** vastly outperform the other method that doesn't use additional information
- **TFq** outperform 2 of the 3 methods that use additional information

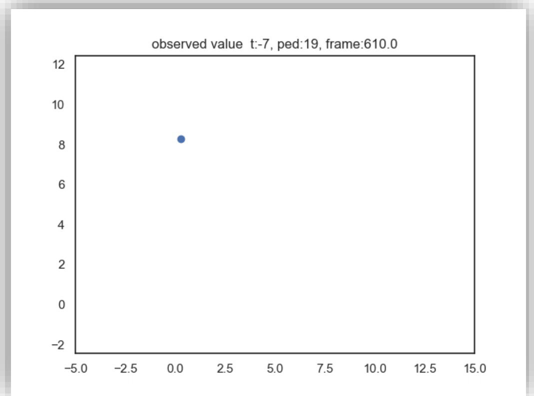
	LSTM-based				TF-based
	Individual	Social		Soc.+ map	Ind.
	S-GAN-ind [10]	S-GAN [10]	Trajectron++ [22]	Soc-BIGAT [14]	TF_q (ours)
ETH	0.81/1.52	*0.87/1.62*	*0.43/0.86*	*0.69/1.29*	**0.61 / 1.12**
Hotel	0.72/1.61	*0.67/1.37*	*0.12/0.19*	*0.49/1.01*	**0.18 / 0.30**
UCY	0.60/1.26	*0.76/1.52*	*0.22/0.43*	*0.55/1.32*	**0.35 / 0.65**
Zara1	0.34/0.69	*0.35/0.68*	*0.17/0.32*	*0.30/0.62*	**0.22 / 0.38**
Zara2	0.42/0.84	*0.42/0.84*	*0.12/0.25*	*0.36/0.75*	**0.17 / 0.32**
Avg	0.58/1.18	*0.61/1.21*	*0.20/0.39*	*0.48/1.00*	**0.31 / 0.55**

Summary

TF



TFq



In the paper:

- Additional tests with the BERT model
- In depth comparison of LSTM vs Transformer
- Ablation studies
  - Longer forecasting horizons
  - Prediction with missing observations

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

IIAI Inception Institute of Artificial Intelligence

SAPIENZA UNIVERSITÀ DI ROMA

THANK YOU

Code is available on github!