Karlsruhe Institute of Technology

UNIVERSIDAD DE GRANADA

Institute for Industrial Production

Applying transformer to imputation of multivariate energy time series data

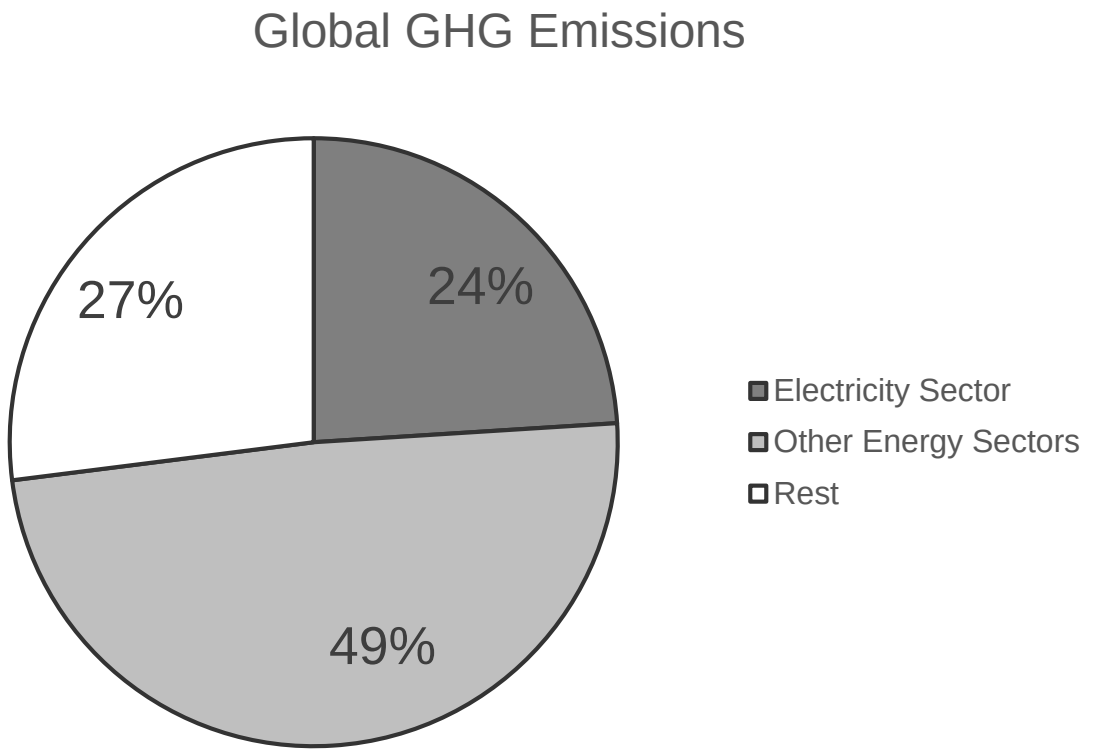
Hasan Ümitcan Yilmaz, Max Kleinebrahm, Christopher Bülte, Juan Gómez Romero

Chair of Energy Economics (Prof Fichtner)
Research Group 'Energy Policy'

Motivation

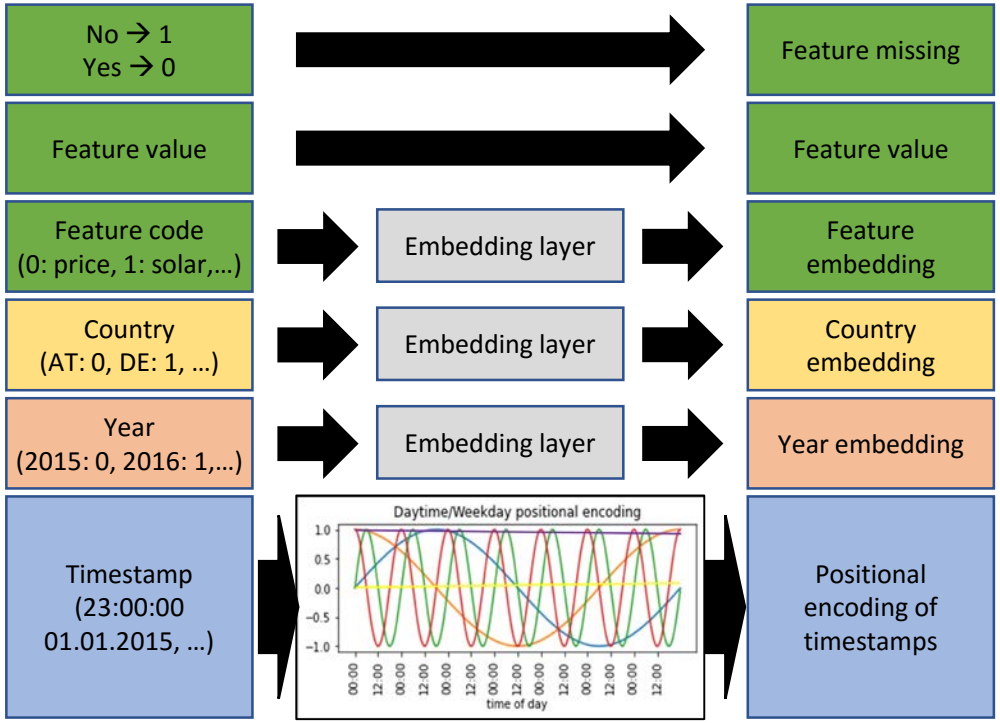
- Global greenhouse gas (GHG) emission
  - 24% electricity sector
  - 49% other energy sectors
- Sector coupling: use of renewable energy sources in other sectors
  - → E.g. Electric cars, heat pumps
- Energy data has many data lacks
- Quality of the data matters in many analyses
- Solution:

  Application of new innovative methods for data imputation: **Transformer**



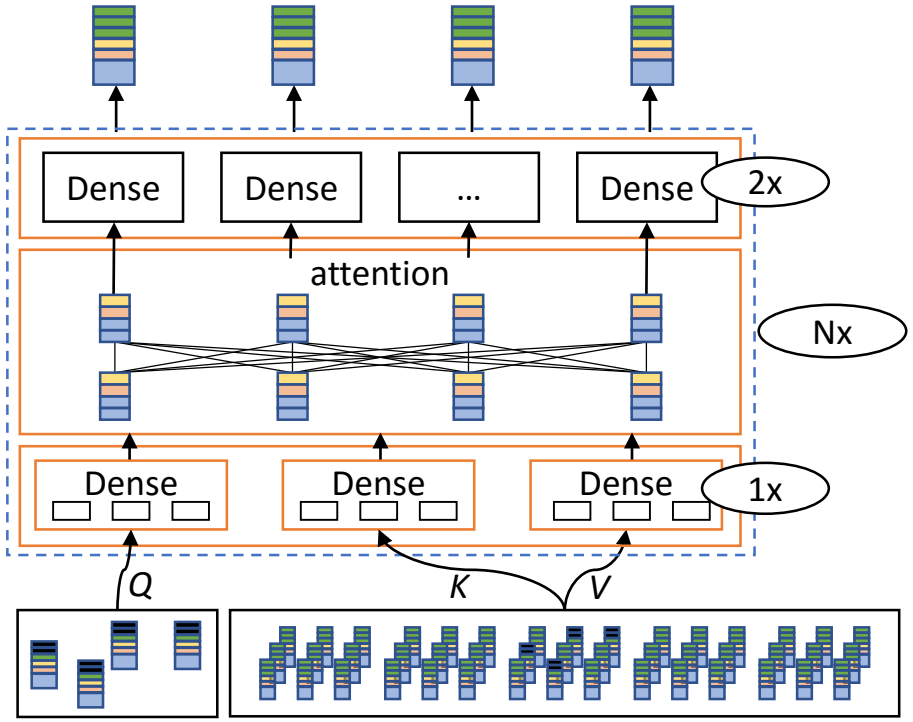
Data processing

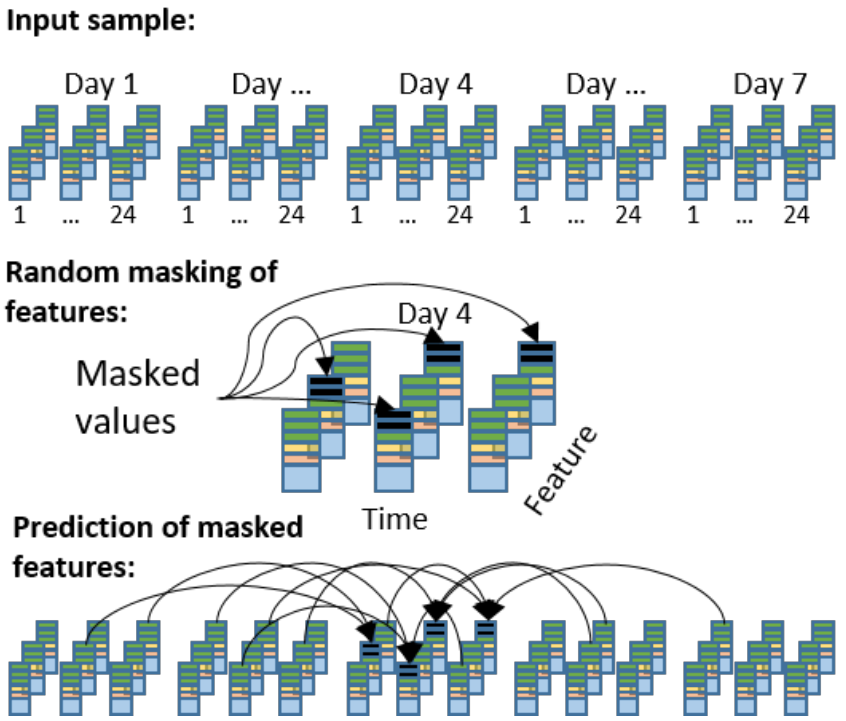
Year	Country	Hour	S_1	S_2	...	S_n
Y_1	C_1	H_1	.	.	.	.
		H_2	.	.	.	.
		...	...	...	...	...
		H_{8760}	.	.	.	.
	C_2	H_1	.	.	.	.
		H_2	.	.	.	.
		...	...	...	...	...
		H_{8760}	.	.	.	.



- Aggregated electricity prices, production and electricity demand data for the European countries
- Embedding layer: feature code, country and year
- Positional encoding: data on position of the information in the sequence

Transformer model

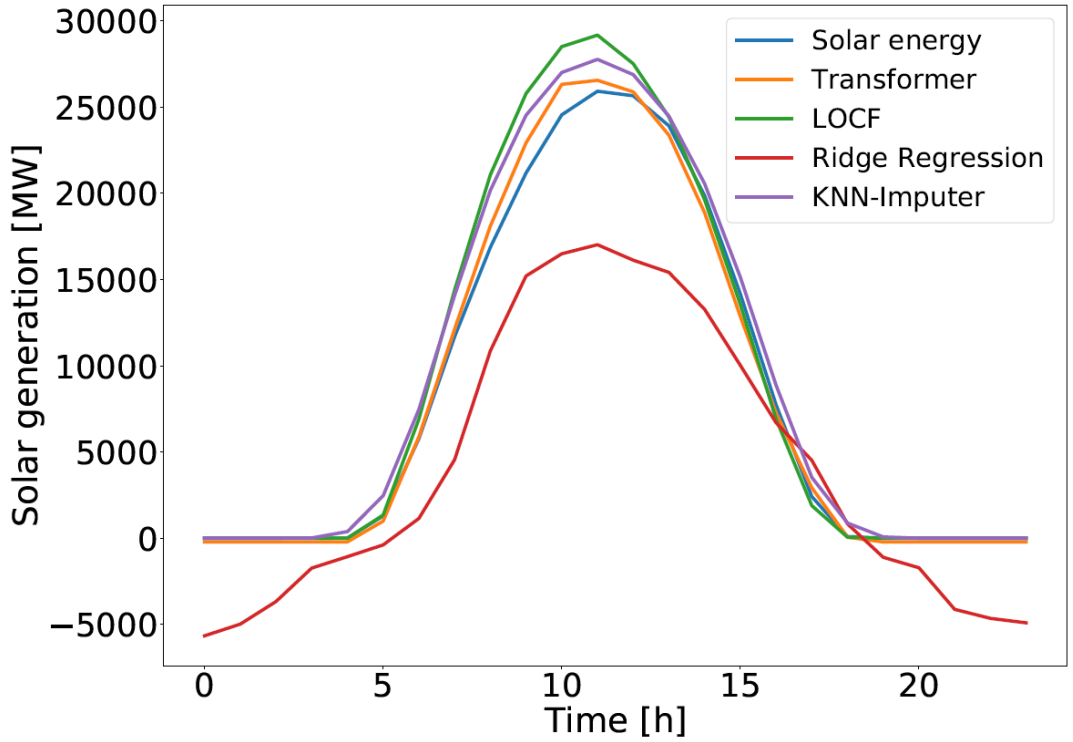




- Training process: mask x data entries of the middle day of the week
- Decomposition mechanism is necessary.

Results

- Only with the data for Germany
- Simplified masking
- Methods for comparison:
  - Last observation carried forward (LOCF)
  - K-Nearest neighbors imputer (KNN-Imputer)
  - Ridge regression



Conclusion:

- Transformer outperforms all other approaches.

	MSE
Transformer	0.0585
LOCF	0.1133
Ridge Regression	0.1827
KNN-Imputer	0.0629

Thank you for the attention!