Self-supervised Learning for Semi-supervised Time Series Classification

Shayan Jawed[✉], Josif Grabocka, and Lars Schmidt-Thieme

Information Systems and Machine Learning Lab, University of Hildesheim, Hildesheim, Germany
{shayan,josif,schmidt-thieme}@ismll.uni-hildesheim.de

Presented by Christer Rustand

Content

- Motivation
- Problem description
- Related work
- Method
- Experimental settings and results
- Conclusion
- My thoughts

Motivation

- High cost to label large datasets
- Gathering large amounts of time series data is often trivial
- So, less labeling * large datasets = winning

Problem description

- Accurately predict classes of time series classification problems where there are few labeled samples using deep learning

Related work

Seminal work

- Nearest neighbor classifiers(with meta-feature distance)
- Clustering
- Graphs(distance functions and label propagation)
- Shapelets

Recent work

- Tasks inherent in the data itself

Most related work

- Multi-task self-supervised network

Method

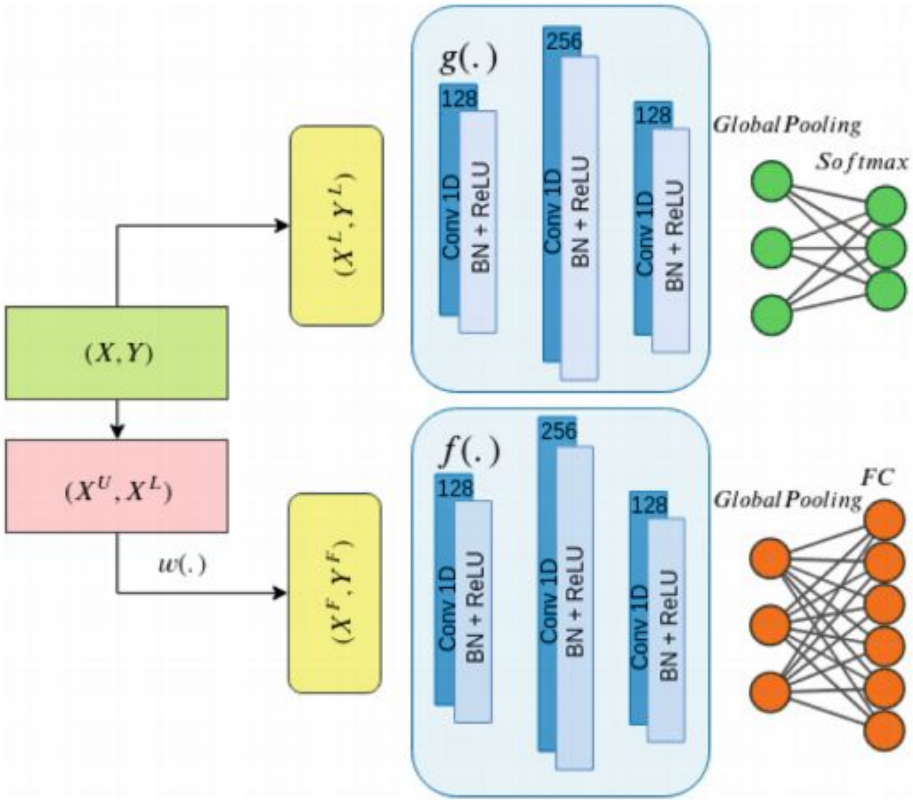
Data preparation

Splitting univariate time series dataset X;

- XL -> Labeled set
- XU -> Unlabeled set

Sliding window function w(.);

- Made up by stride s and horizon h
- Builds forecasting samples XF



Method

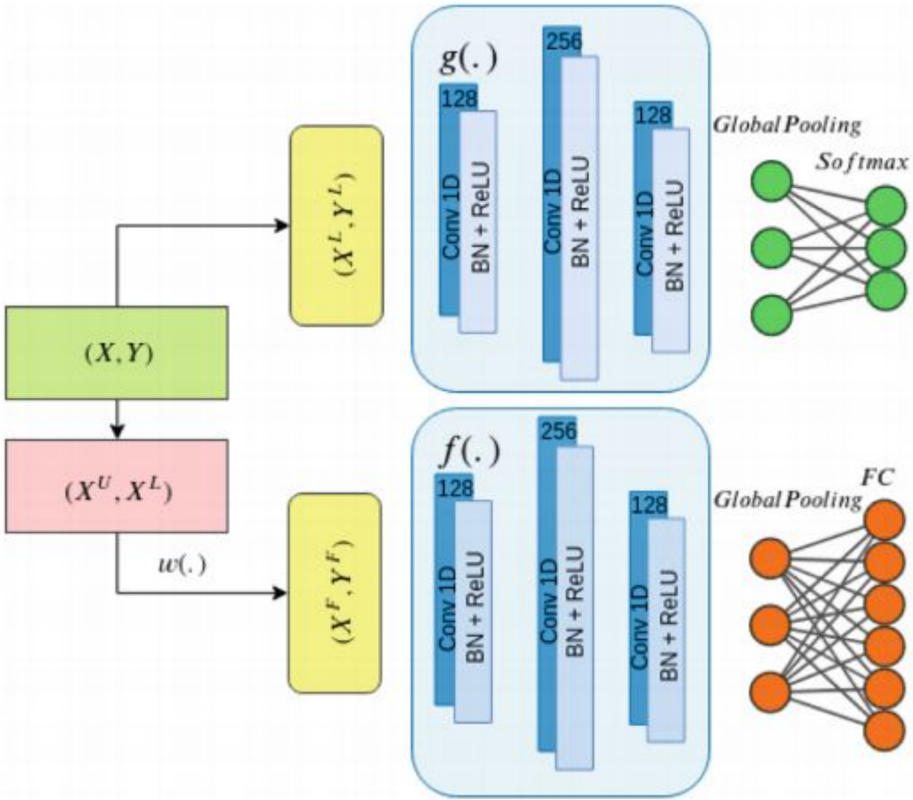
Objective

Forecasting function f(.)

Classification function g(.)

YF = f(XF)

YL = g(XL)



Method

Loss functions

Forecasting loss function Lf(.)

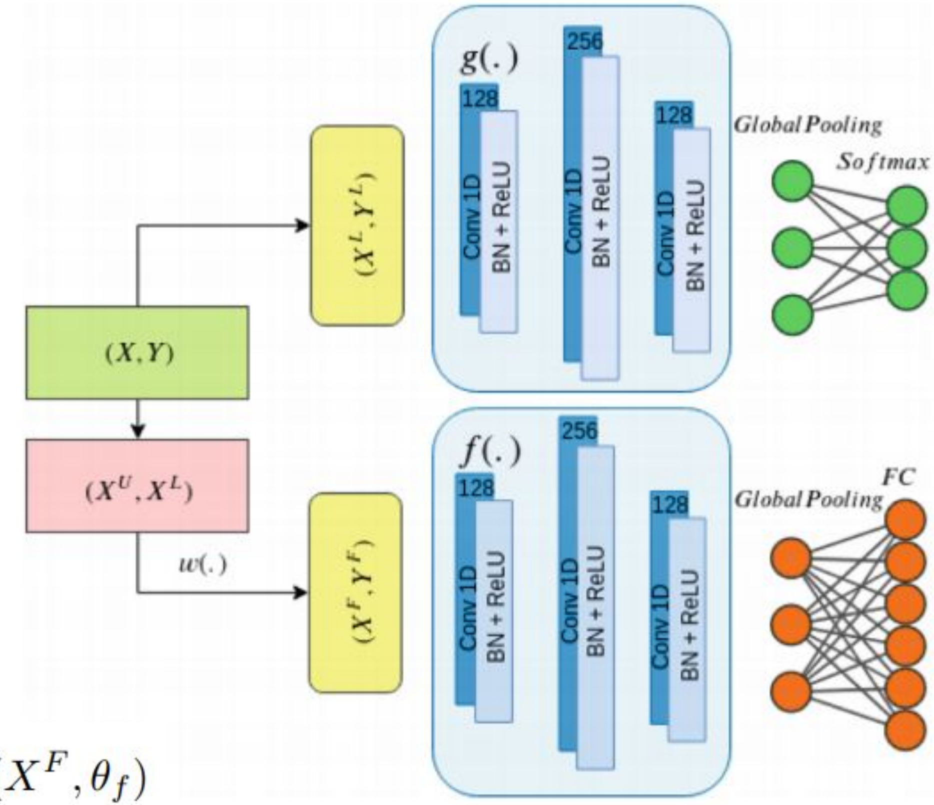
$$L_f(X^F, \theta_f) = \frac{1}{n \times m \times h} \sum_{i}^{n} \sum_{j}^{m} \sum_{t}^{h} (y_{jt}^{i} - \hat{y}_{jt}^{i})^2$$

Classification loss function Lc(.)

$$L_c(X^L, \theta_c) = -\frac{1}{l} \sum_{i}^{l} log \left(\frac{e^{\hat{y}_{i=c}}}{\sum_{j}^{C} e^{\hat{y}_i}} \right)$$

Combined loss function

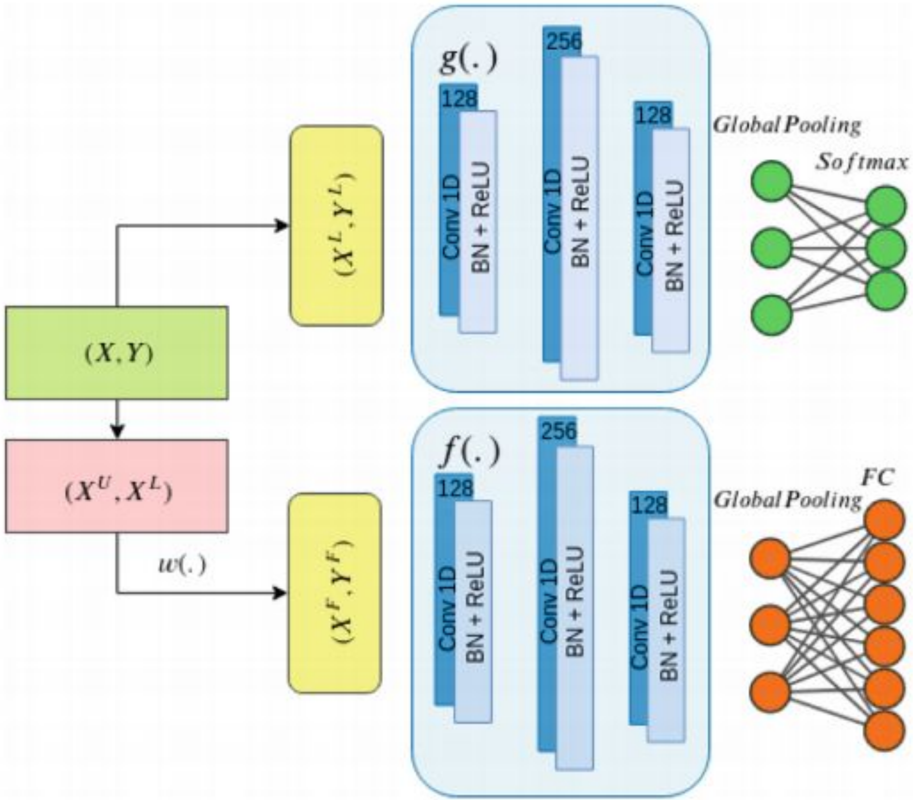
$$L_{MTL}(X^F, \theta_f, X^L, \theta_c) = L_c(X^L, \theta_c) + \lambda L_f(X^F, \theta_f)$$



Method

Further elaboration

- Forecasting as an auxiliary task
- Horizon and stride
- Multi-task learning
- Key challenges
- Intuition?



Experimental settings and results

Baselines

- Wei’s method
- DTW-D
- SUCCESS
- Xu’s method
- Bag-of-words
- SSSL
- Base
- Π-model
- Transfer learning

Results

	Coffee	CBF	ECG	Face-Four	OSULf	Italy-Power	Light.2	Light.7	Gun-Point	Trace	Word-Syn	Olive-Oil	Star-Light
#	56	930	200	112	442	1096	121	143	200	200	905	60	9236
C	2	3	2	4	6	2	2	7	2	4	25	4	3
T	286	128	96	350	427	24	637	319	150	275	270	570	1024

Datasets	Results verbatim from table in [17]						Proposed			
	Wei.	DTW-D	SUC.	Xu.	BoW	SSSL	Base	Π	Tr.	MTL
Coffee	0.571	0.601	0.632	0.588	0.620	0.792	**1.0**	**1.0**	**1.0**	**1.0**
CBF	0.995	0.833	0.997	0.921	0.873	**1.0**	**1.0**	**1.0**	0.784	**1.0**
cre ECG	0.763	0.953	0.775	0.819	0.955	0.793	0.9	0.875	0.9	**0.975**
FaceFour	0.818	0.782	0.800	0.833	0.744	0.851	0.913	0.913	0.739	**0.957**
OSULf	0.468	0.701	0.534	0.642	0.685	0.835	0.977	0.977	0.460	**0.978**
ItalyPower	0.934	0.664	0.924	0.772	0.813	0.941	0.986	0.986	0.959	**0.991**
Light.2	0.658	0.641	0.683	0.698	0.721	0.813	0.92	0.84	0.88	**0.92**
Light.7	0.464	0.503	0.471	0.511	0.677	0.796	0.758	0.689	0.482	**0.828**
GunPoint	0.925	0.711	0.955	0.729	0.925	0.824	**1.0**	**1.0**	0.825	**1.0**
Trace	0.950	0.801	**1.0**	0.788	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
WordSyn	0.590	0.863	0.618	0.639	0.795	**0.875**	0.497	0.491	0.342	0.519
OliveOil	0.633	0.732	0.617	0.639	0.766	0.776	0.916	**1.0**	0.833	**1.0**
StarLight	0.860	0.743	0.800	0.755	0.851	0.872	0.982	0.983	**1.0**	0.991

Settings used

Datasets	s: 0.05		0.1		0.2	
	h: 0.1	0.2	0.1	0.2	0.1	0.2
Coffee	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
CBF	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
ECG	0.950	0.9	0.925	0.9	**0.975**	0.875
FaceFr.	0.913	0.913	0.870	**0.957**	**0.957**	0.913
OSULf.	0.966	0.966	**0.978**	0.955	**0.978**	0.955
ItalyPower	0.986	**0.991**	0.986	0.986	0.982	**0.991**
Light.2	0.840	**0.920**	0.840	0.880	0.880	0.880
Light.7	**0.828**	**0.828**	0.759	0.759	0.793	0.759
GunPoint	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Trace	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
WordSyn.	0.497	**0.519**	0.508	0.497	0.503	0.508
OliveOil	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
StarLight	0.983	0.983	0.99	0.97	0.983	**0.991**

Conclusion

- Novel self-supervised task
- More accurate than a majority of baselines on benchmark datasets
- Multivariate time series next
- Github repo available for reproducibility

My thoughts

- Liked the concept
- Not too hard to read
- Maybe relevant to my thesis