An Efficient Empirical Solver for Localized Multiple Kernel Learning via DNNs

Ziming Zhang

Department of Electrical & Computer Engineering

Worcester Polytechnic Institute

Problem Definition

- Multiple kernel learning (MKL)

$$f_{Simple}(z) = \sum_i \alpha_i y_i \left[\sum_m \beta_m \mathbf{K}_m(z, x_i) \right]$$

- Localized MKL (LMKL)

$$f_{Loc}(z) = \sum_i \alpha_i y_i \left[\sum_m \eta_m(z) \mathbf{K}_m(z, x_i) \eta_m(x_i) \right] + b$$

gating function

Related Work

- LMKL
  - In the literature, the classifier parameters are shared among all the data samples, with some prior on the gating function that may be wrong empirically
- Large-scale MKL
  - To the best of our knowledge, no such approaches are proposed for LMKL
- Optimization
  - Many approaches (e.g. SILP-MKL, SPG-GMKL, etc.) were proposed from this perspective
- **In contrast**, we propose a deep learning solution
  - using an attentional network to approximate the unknown gating function
  - much faster training speed and much smaller memory footprint for large-scale LMKL with better accuracy

Our Approach

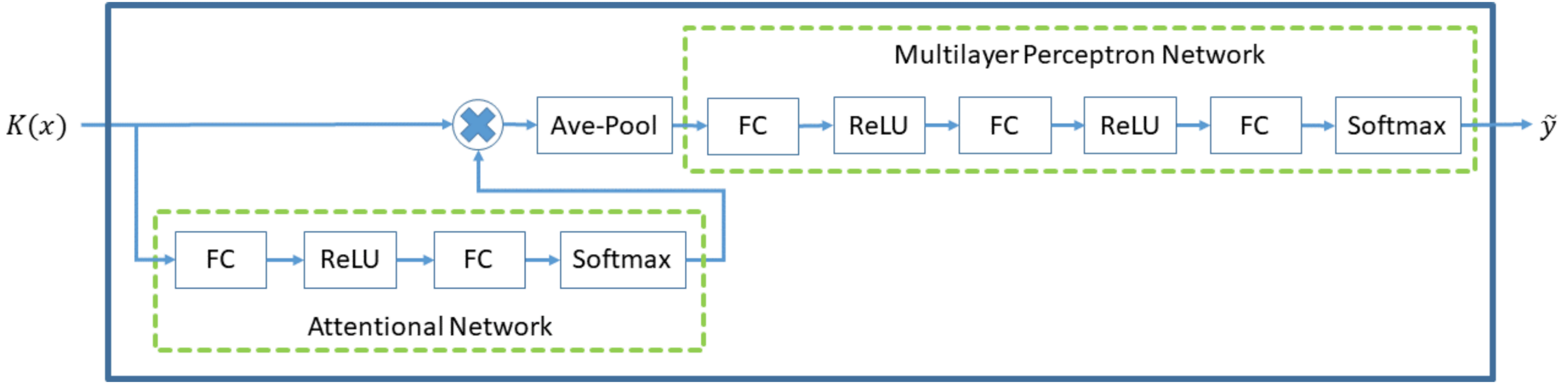
- Motivation

$$\sum_m \eta_m(z)\mathbf{K}_m(z, x_i)\eta_m(x_i) \;=\; \sum_m \mathbf{K}_m(z, x_i)\boxed{\mathbf{K}_{\eta_m}(z, x_i)}$$

$$f(z; \omega, \pi) = g\left(\sum_{m=1}^{M} \boxed{h\Big(\mathbf{K}(z); \omega\Big)} \otimes \mathbf{K}(z); \pi\right)$$

Cont.

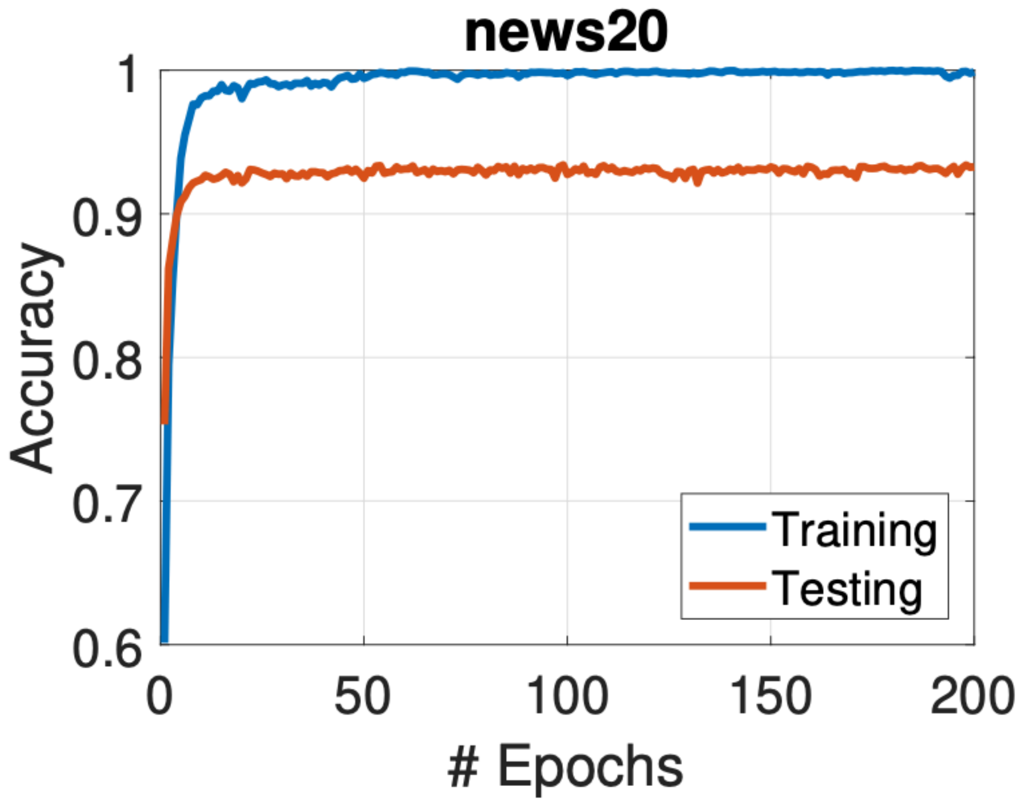
- LMKL-Net

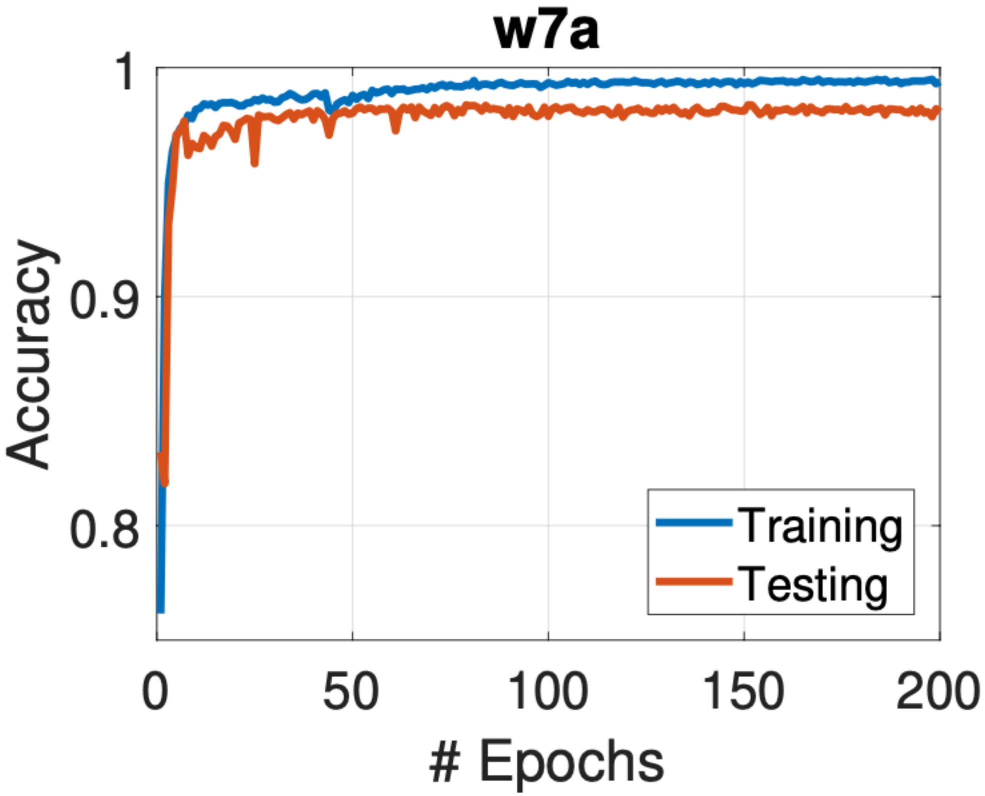


Some Results

	adult-8	news20	phishing	rcv1	real-sim	w7a	average
UNIFORM	81.94	93.33	46.16	96.37	96.56	90.37	84.12
SPG-GMKL	84.13	90.27	95.26	95.57	92.21	97.05	92.42
EasyMKL	84.57	91.32	96.74	95.08	94.23	97.95	93.32
LMKL	78.09	**95.52**	52.04	96.76	**97.09**	97.75	86.21
Lp-MKL	76.33	-	-	-	-	97.05	-
OBSCURE	84.22	94.68	97.25	96.55	96.94	98.50	94.69
Ours (with 10 kernels)	84.62	93.53	98.17	96.71	96.42	98.74	94.70
Ours (with 1K kernels)	**85.43**	94.06	**98.70**	**97.32**	96.25	**98.76**	**95.09**

Cont.


news20
Accuracy
1
0.9
0.8
0.7
0.6
0
50
100
150
200
Epochs
Training
Testing


w7a
Accuracy
1
0.9
0.8
0
50
100
150
200
Epochs
Training
Testing

Conclusion

- LMKL-Net: A deep learning solution for LMKL
  - Using an attentional network to approximate the unknown gating function
  - Suitable for large-scale learning
  - Faster training speed, smaller memory footprint, and better accuracy
- For more details, please refer to our paper