

Adversarial Multi-task Learning for Text Classification

Yunhan Bai
2019.4.12

OUTLINE

1.Motivation

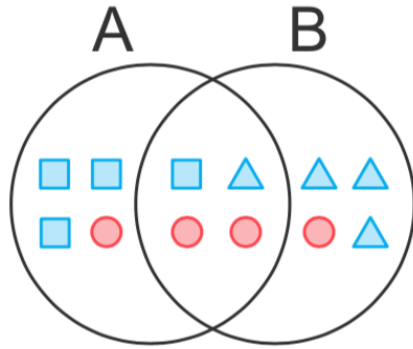
2.Adversarial Training Strategy

3.Result

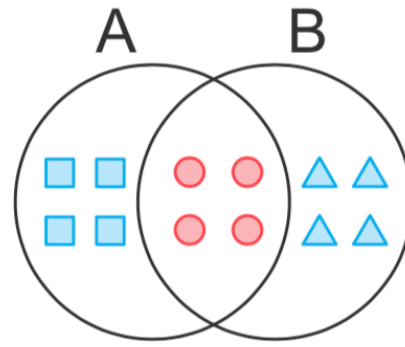
1. Motivation

- MTL has been proved to has potential to improve ml task performance by training multi-tasks together.
- The underlying detail of MTL is parameter sharing, which is similar to transformer learning.
- Which part of parameter should be shared and vice versa ?

1. Motivation



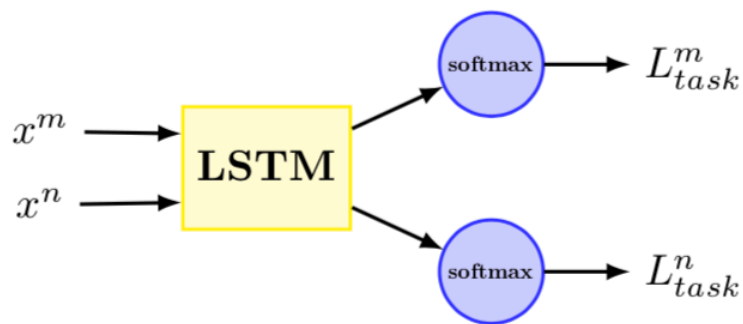
(a) Shared-Private Model



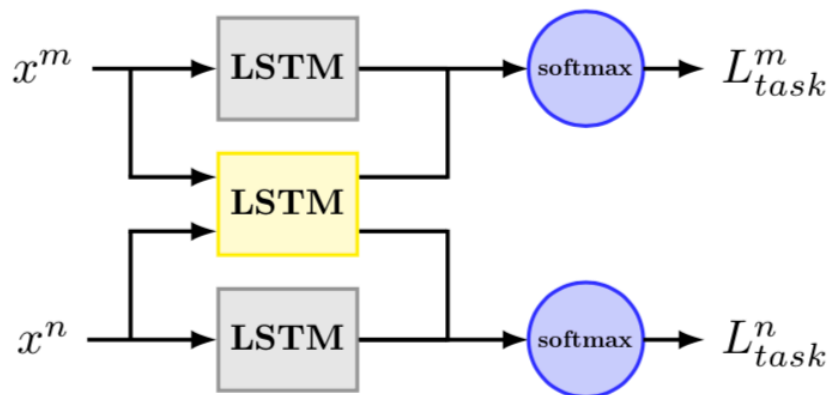
(b) Adversarial Shared-Private Model

- The key point or prior is : shared parameters(or feature map) should be task irrelevant.

2. Adversarial Training Strategy



(a) Fully Shared Model (FS-MTL)

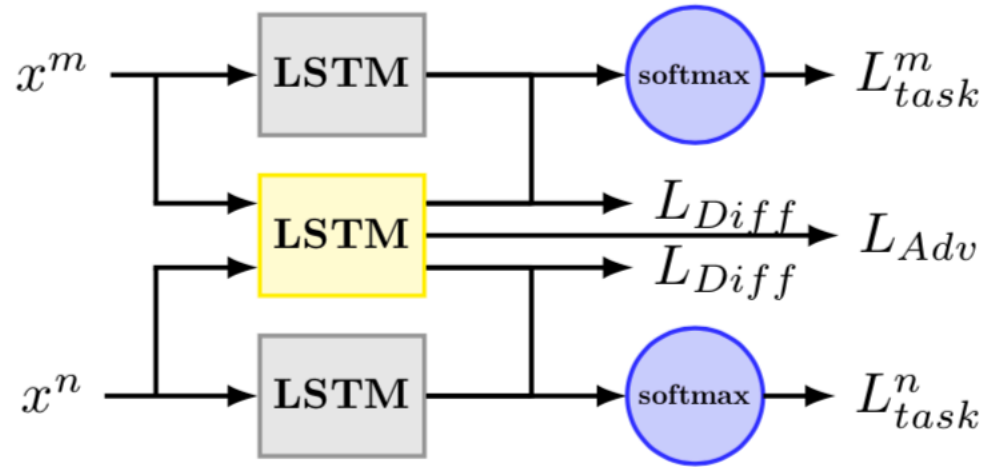


(b) Shared-Private Model (SP-MTL)

$$L_{Task} = \sum_{k=1}^K \alpha_k L(\hat{y}^{(k)}, y^{(k)})$$

2. Adversarial Training Strategy

- Adversarial Network



$$D(\mathbf{s}_T^k, \theta_D) = \text{softmax}(\mathbf{b} + \mathbf{U}\mathbf{s}_T^k)$$

$$L_{Adv} = \min_{\theta_s} \left(\lambda \max_{\theta_D} \left(\sum_{k=1}^K \sum_{i=1}^{N_k} d_i^k \log[D(E(\mathbf{x}^k))] \right) \right)$$

2. Adversarial Training Strategy

- Orthogonality Constraints

$$L_{\text{diff}} = \sum_{k=1}^K \left\| \mathbf{S}^{k\top} \mathbf{H}^k \right\|_F^2 ,$$

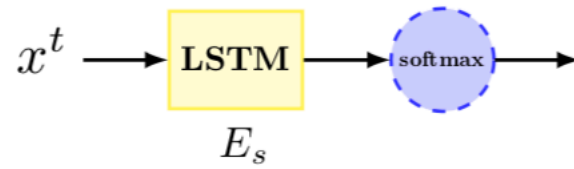
- Total Loss

$$L = L_{Task} + \lambda L_{Adv} + \gamma L_{Diff}$$

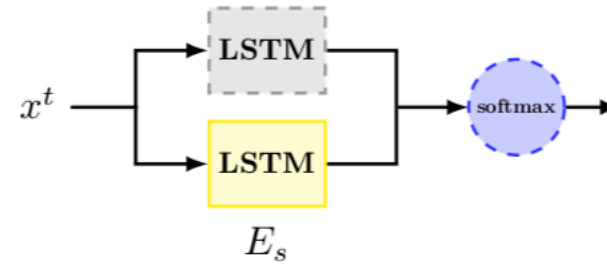
3. Result

Task	Single Task				Multiple Tasks				
	LSTM	BiLSTM	sLSTM	Avg.	MT-DNN	MT-CNN	FS-MTL	SP-MTL	ASP-MTL
Books	20.5	19.0	18.0	19.2	17.8 _(-1.4)	15.5 _(-3.7)	17.5 _(-1.7)	18.8 _(-0.4)	16.0 _(-3.2)
Electronics	19.5	21.5	23.3	21.4	18.3 _(-3.1)	16.8 _(-4.6)	14.3 _(-7.1)	15.3 _(-6.1)	13.2 _(-8.2)
DVD	18.3	19.5	22.0	19.9	15.8 _(-4.1)	16.0 _(-3.9)	16.5 _(-3.4)	16.0 _(-3.9)	14.5 _(-5.4)
Kitchen	22.0	18.8	19.5	20.1	19.3 _(-0.8)	16.8 _(-3.3)	14.0 _(-6.1)	14.8 _(-5.3)	13.8 _(-6.3)
Apparel	16.8	14.0	16.3	15.7	15.0 _(-0.7)	16.3 _(+0.6)	15.5 _(-0.2)	13.5 _(-2.2)	13.0 _(-2.7)
Camera	14.8	14.0	15.0	14.6	13.8 _(-0.8)	14.0 _(-0.6)	13.5 _(-1.1)	12.0 _(-2.6)	10.8 _(-3.8)
Health	15.5	21.3	16.5	17.8	14.3 _(-3.5)	12.8 _(-5.0)	12.0 _(-5.8)	12.8 _(-5.0)	11.8 _(-6.0)
Music	23.3	22.8	23.0	23.0	15.3 _(-7.7)	16.3 _(-6.7)	18.8 _(-4.2)	17.0 _(-6.0)	17.5 _(-5.5)
Toys	16.8	15.3	16.8	16.3	12.3 _(-4.0)	10.8 _(-5.5)	15.5 _(-0.8)	14.8 _(-1.5)	12.0 _(-4.3)
Video	18.5	16.3	16.3	17.0	15.0 _(-2.0)	18.5 _(+1.5)	16.3 _(-0.7)	16.8 _(-0.2)	15.5 _(-1.5)
Baby	15.3	16.5	15.8	15.9	12.0 _(-3.9)	12.3 _(-3.6)	12.0 _(-3.9)	13.3 _(-2.6)	11.8 _(-4.1)
Magazines	10.8	8.5	12.3	10.5	10.5 _(+0.0)	12.3 _(+1.8)	7.5 _(-3.0)	8.0 _(-2.5)	7.8 _(-2.7)
Software	15.3	14.3	14.5	14.7	14.3 _(-0.4)	13.5 _(-1.2)	13.8 _(-0.9)	13.0 _(-1.7)	12.8 _(-1.9)
Sports	18.3	16.0	17.5	17.3	16.8 _(-0.5)	16.0 _(-1.3)	14.5 _(-2.8)	12.8 _(-4.5)	14.3 _(-3.0)
IMDB	18.3	15.0	18.5	17.3	16.8 _(-0.5)	13.8 _(-3.5)	17.5 _(+0.2)	15.3 _(-2.0)	14.5 _(-2.8)
MR	27.3	25.3	28.0	26.9	24.5 _(-2.4)	25.5 _(-1.4)	25.3 _(-1.6)	24.0 _(-2.9)	23.3 _(-3.6)
AVG	18.2	17.4	18.3	18.0	15.7 _(-2.2)	15.5 _(-2.5)	15.3 _(-2.7)	14.9 _(-3.1)	13.9 _(-4.1)

3. Result



(a) Single Channel

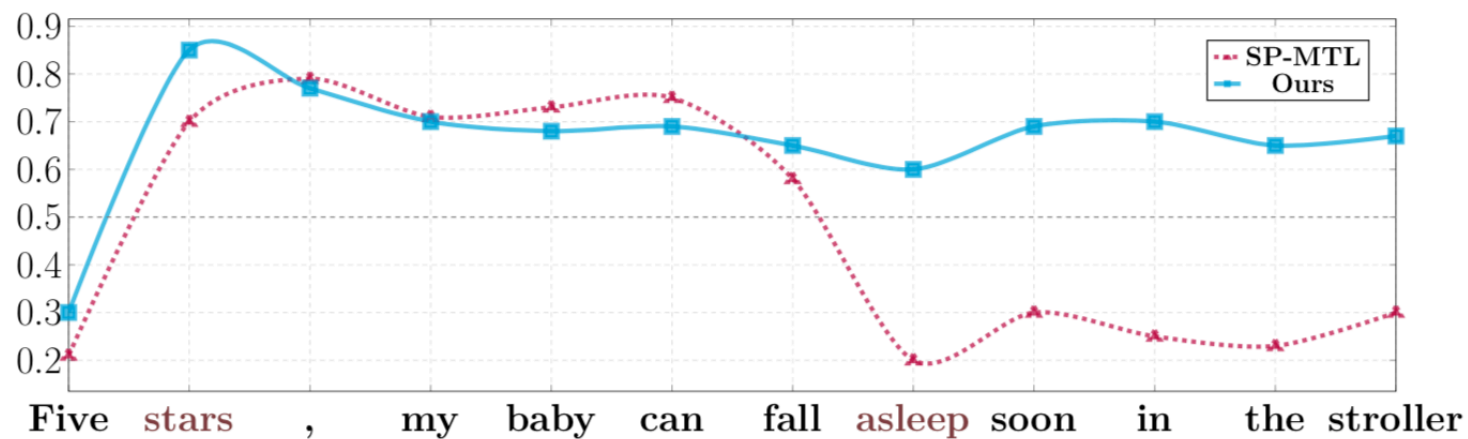


(b) Bi-Channel

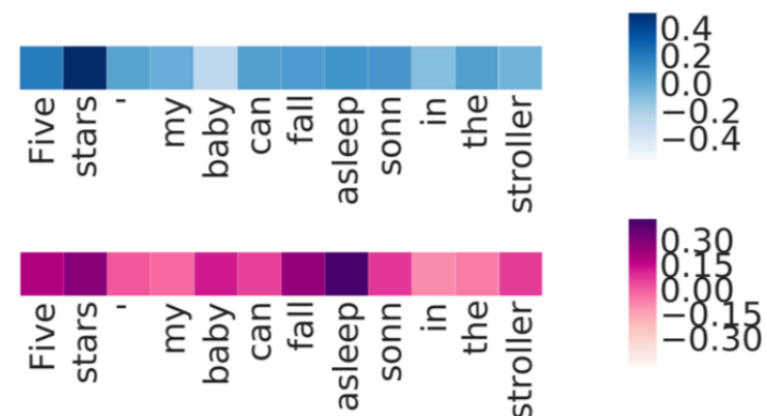
3. Result

Source Tasks	Single Task				Transfer Models			
	LSTM	BiLSTM	sLSTM	Avg.	SP-MTL-SC	SP-MTL-BC	ASP-MTL-SC	ASP-MTL-BC
ϕ (Books)	20.5	19.0	18.0	19.2	17.8 _(-1.4)	16.3 _(-2.9)	16.8 _(-2.4)	16.3 _(-2.9)
ϕ (Electronics)	19.5	21.5	23.3	21.4	15.3 _(-6.1)	14.8 _(-6.6)	17.8 _(-3.6)	16.8 _(-4.6)
ϕ (DVD)	18.3	19.5	22.0	19.9	14.8 _(-5.1)	15.5 _(-4.4)	14.5 _(-5.4)	14.3 _(-5.6)
ϕ (Kitchen)	22.0	18.8	19.5	20.1	15.0 _(-5.1)	16.3 _(-3.8)	16.3 _(-3.8)	15.0 _(-5.1)
ϕ (Apparel)	16.8	14.0	16.3	15.7	14.8 _(-0.9)	12.0 _(-3.7)	12.5 _(-3.2)	13.8 _(-1.9)
ϕ (Camera)	14.8	14.0	15.0	14.6	13.3 _(-1.3)	12.5 _(-2.1)	11.8 _(-2.8)	10.3 _(-4.3)
ϕ (Health)	15.5	21.3	16.5	17.8	14.5 _(-3.3)	14.3 _(-3.5)	12.3 _(-5.5)	13.5 _(-4.3)
ϕ (Music)	23.3	22.8	23.0	23.0	20.0 _(-3.0)	17.8 _(-5.2)	17.5 _(-5.5)	18.3 _(-4.7)
ϕ (Toys)	16.8	15.3	16.8	16.3	13.8 _(-2.5)	12.5 _(-3.8)	13.0 _(-3.3)	11.8 _(-4.5)
ϕ (Video)	18.5	16.3	16.3	17.0	14.3 _(-2.7)	15.0 _(-2.0)	14.8 _(-2.2)	14.8 _(-2.2)
ϕ (Baby)	15.3	16.5	15.8	15.9	16.5 _(+0.6)	16.8 _(+0.9)	13.5 _(-2.4)	12.0 _(-3.9)
ϕ (Magazines)	10.8	8.5	12.3	10.5	10.5 _(+0.0)	10.3 _(-0.2)	8.8 _(-1.7)	9.5 _(-1.0)
ϕ (Software)	15.3	14.3	14.5	14.7	13.0 _(-1.7)	12.8 _(-1.9)	14.5 _(-0.2)	11.8 _(-2.9)
ϕ (Sports)	18.3	16.0	17.5	17.3	16.3 _(-1.0)	16.3 _(-1.0)	13.3 _(-4.0)	13.5 _(-3.8)
ϕ (IMDB)	18.3	15.0	18.5	17.3	12.8 _(-4.5)	12.8 _(-4.5)	12.5 _(-4.8)	13.3 _(-4.0)
ϕ (MR)	27.3	25.3	28.0	26.9	26.0 _(-0.9)	26.5 _(-0.4)	24.8 _(-2.1)	23.5 _(-3.4)
AVG	18.2	17.4	18.3	18.0	15.6 _(-2.4)	15.2 _(-2.8)	14.7 _(-3.3)	14.3 _(-3.7)

3. Result



(a) Predicted Sentiment Score by Two Models



(b) Behaviours of Neuron h_{18}^s and h_{21}^s



Thank You

Q&A