ICML 2020

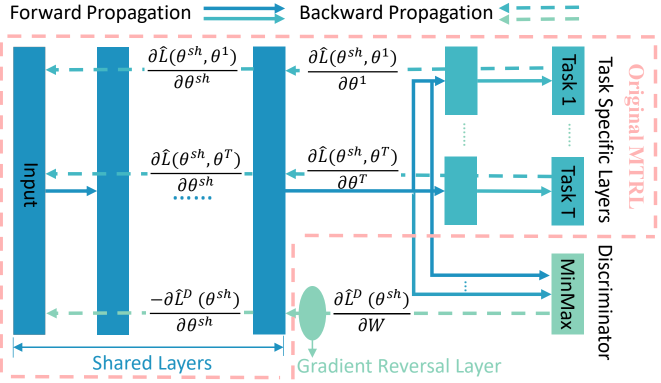
Adaptive Adversarial Multi-task Representation Learning

Yuren Mao[1] Weiwei Liu[2] Xuemin Lin[1]

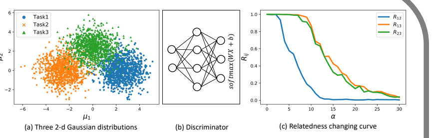
1. University of New South Wales, Australia.
2. Wuhan University, China.

Overview: Adaptive AMTRL (Adversarial Multi-task Representation Learning)

Algorithm



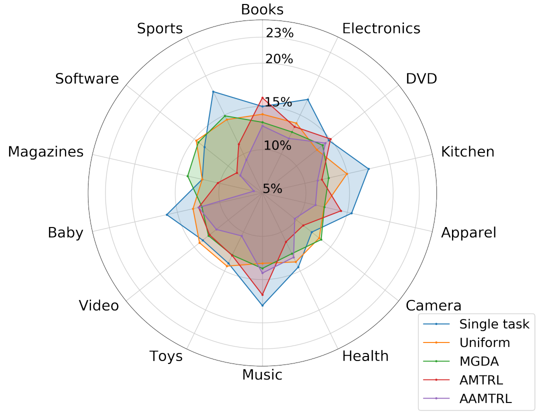
AMTRL



Task Relatedness for AMTRL

Adaptive AMTRL

- Augmented Lagrangian
- Relatedness based Weighting Strategy



Better Performance

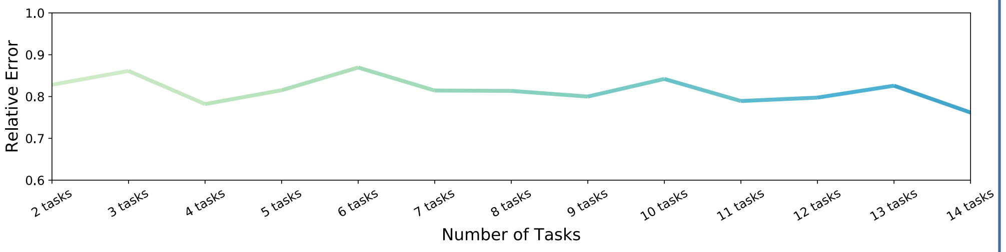
PAC Bound

$$\mathcal{L}_{\mathcal{D}}(h) - \mathcal{L}_{S}(h) \leq \frac{c_1 \rho \ G_a(\mathcal{G}^*(\overline{X}_1))}{n} + \frac{c_2 Q sup_{g \in \mathcal{G}^*} \|g(\overline{X}_1)\|}{\sqrt{n}} + \sqrt{\frac{9 ln(2/\delta)}{2nT}}$$

Generalization Error

The number of tasks does not matter

Negligible

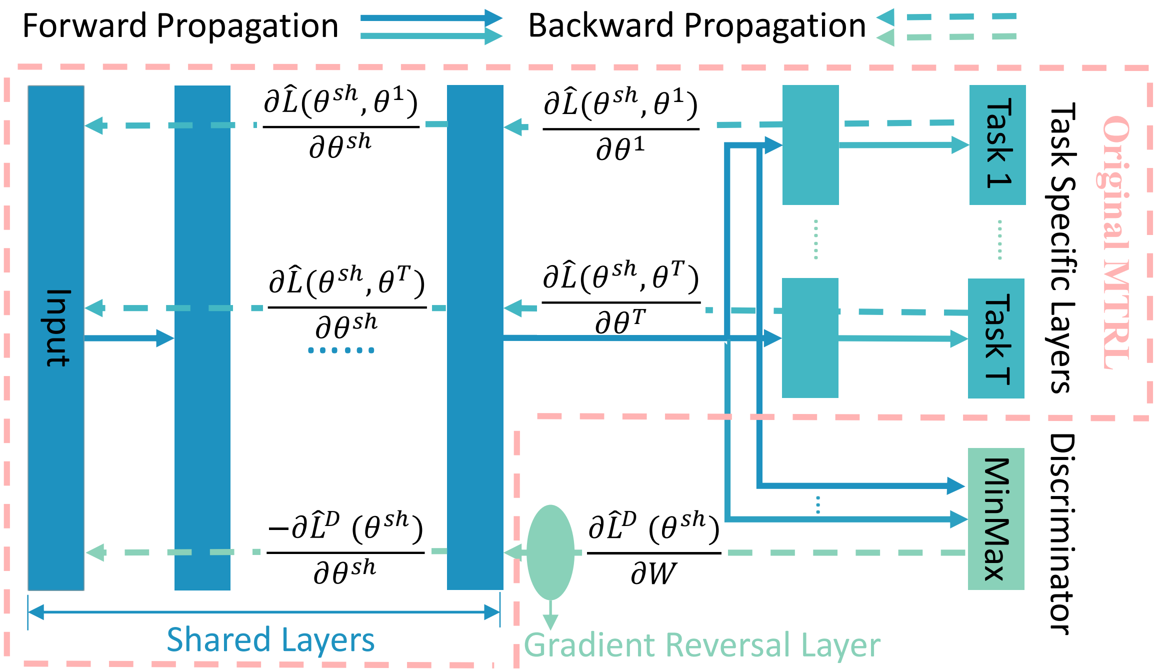


Content

- Adversarial Multi-task Representation Learning (AMTRL)
- Adaptive AMTRL
- PAC Bound and Analysis
- Experiments

Adversarial Multi-task Representation Learning

Adversarial Multi-task Representation Learning (AMTRL) has achieved success in various applications, ranging from sentiment analysis to question answering systems.



$$\min_h L(h,\lambda) = \mathcal{L}_S(h) + \lambda \mathcal{L}^{adv}$$

Empirical loss:

$$\mathcal{L}_S(h) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} l^t(f^t(g(x_i^t)), y_i^t)$$

Loss of the adversarial module:

$$\mathcal{L}^{adv} = \max_{\Phi} \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} e_t \Phi(g(x_i^t))$$

Adaptive AMTRL

Adversarial AMTRL aims to minimize the task-averaged empirical risk and enforce the representation of each task to share an identical distribution. We formulate it as a constraint optimization problem

$$\begin{aligned} \min_{h} \quad & \mathcal{L}_S(h) \\ s.t. \quad & \mathcal{L}^{adv} - c = 0, \end{aligned}$$

and propose to solve the problem with an augmented Lagrangian method.

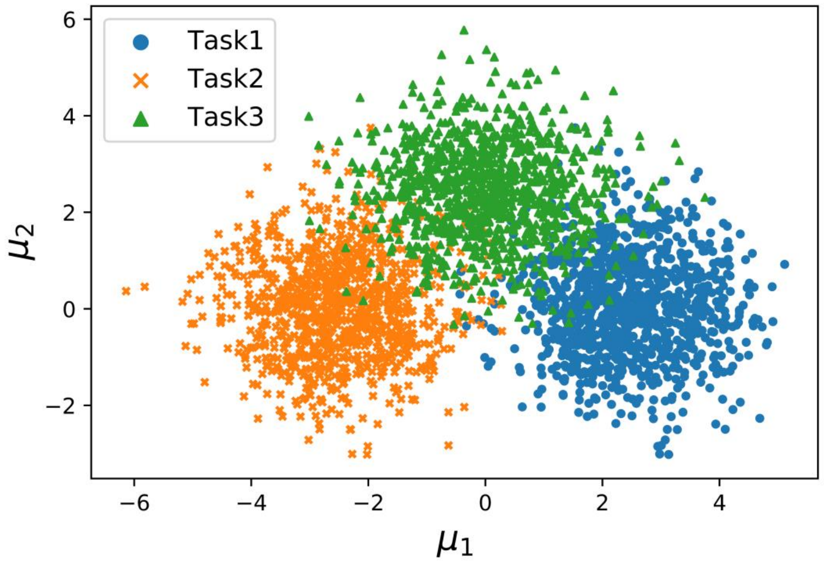
$$\min_{h} \frac{1}{T}\mathcal{L}_S(h) + \lambda(\mathcal{L}^{adv} - c) + \frac{r}{2}(\mathcal{L}^{adv} - c)^2.$$

λ and r updates in the training process.

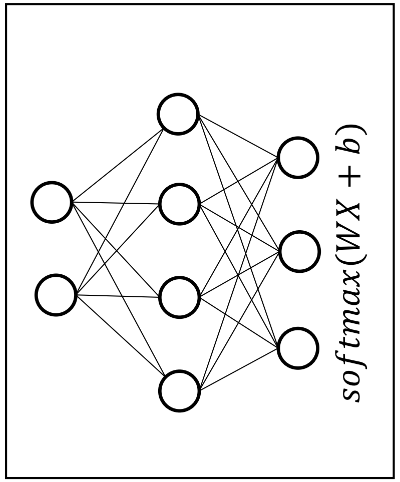
Relatedness for AMTRL

Relatedness between task i and task j: $$R_{ij} = \min\{\frac{\sum_{n=1}^{N} e_j \Phi(g(x_n^i)) + e_i \Phi(g(x_n^j))}{\sum_{n=1}^{N} e_i \Phi(g(x_n^i)) + e_j \Phi(g(x_n^j))}, 1\}$$

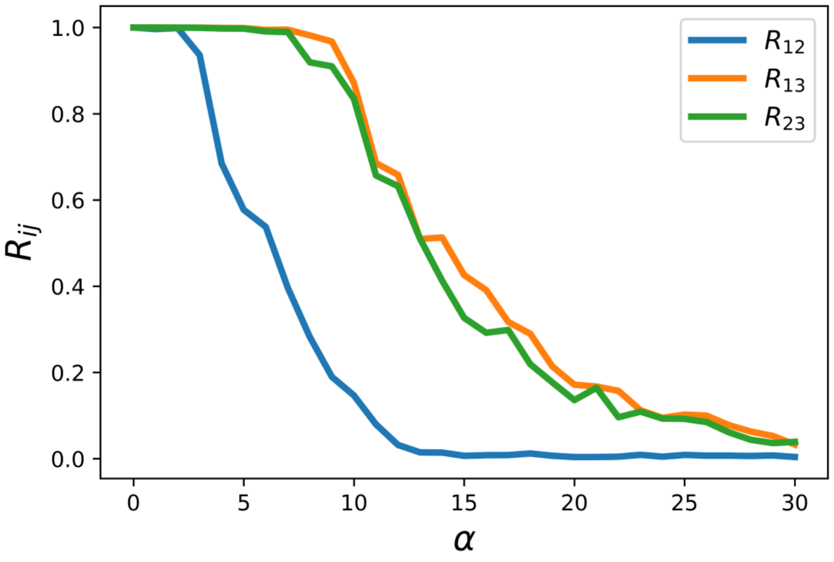
Relatedness matrix: $$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1T} \\ R_{21} & R_{22} & \cdots & R_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ R_{T1} & R_{T2} & \cdots & R_{TT} \end{bmatrix}.$$



(a) Three 2-d Gaussian distributions



(b) Discriminator



(c) Relatedness changing curve

Adaptive AMTRL

In multi-task learning, tasks regularize each other and improve the generalization of some tasks. The weights of each task influences the effect of the regularization. This paper proposes a weighting strategy for AMTRL based on the proposed task relatedness.

$$\mathbf{w} = \frac{1}{\mathbf{1}R\mathbf{1}'}\mathbf{1}R,$$

where 1 is a $1 \times T$ vector of all 1, and R is the relatedness matrix.

Combining the augmented Lagrangian method with the weighting strategy, optimization objective of our adaptive AMTRL method is

$$\min_h \frac{1}{T}\sum_{t=1}^{T} w_t \mathcal{L}_{S_t}(f^t \circ g) + \lambda(\mathcal{L}^{adv} - c) + \frac{r}{2}(\mathcal{L}^{adv} - c)^2.$$

PAC Bound and Analysis

Assume the representation of each task share an identical distribution, we have the following generalization error bound.

$$\mathcal{L}_{\mathcal{D}}(h) - \mathcal{L}_S(h) \leq \frac{c_1 \rho \; G_a(\mathcal{G}^*(\overline{X}_1))}{n} + \frac{c_2 Q sup_{g \in \mathcal{G}^*} \|g(\overline{X}_1)\|}{\sqrt{n}} + \sqrt{\frac{9 ln(2/\delta)}{2nT}}$$

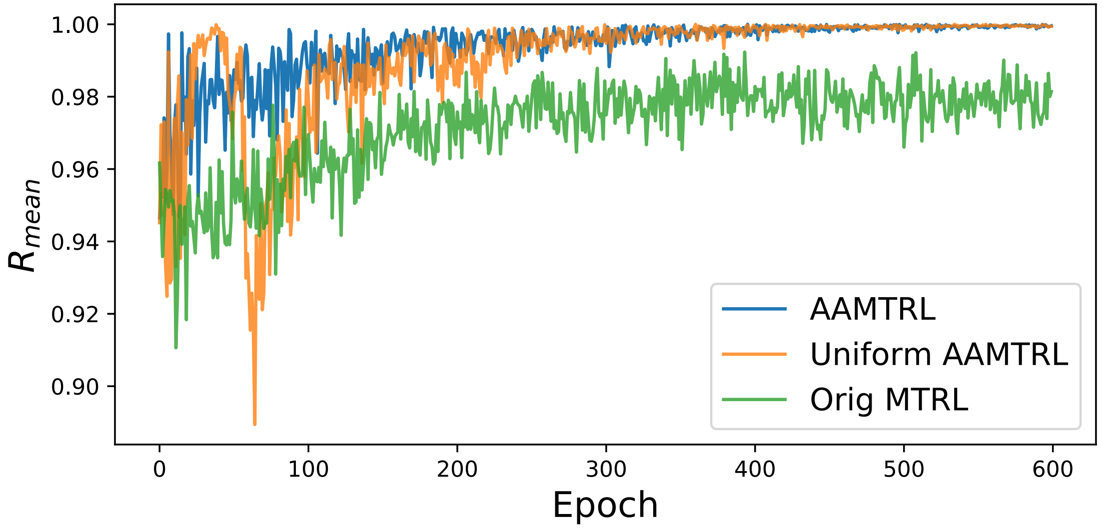
Generalization Error | The number of tasks does not matter | Negligible

- The generalization error bound for AMTRL is tighter than that for MTRL.
- The number of tasks slightly influence the generalization bound of AMTRL.

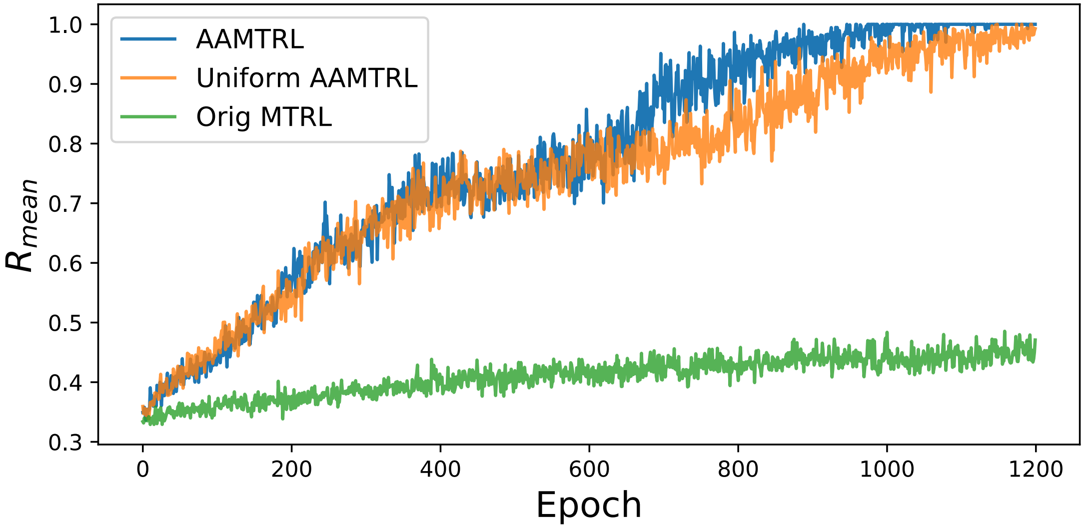
Experiments - *Relatedness Evolution*

Sentiment Analysis and Topic Classification.

Mean of $R_t = \frac{1}{T}\sum_{k=0}^{T} R_{tk}.$



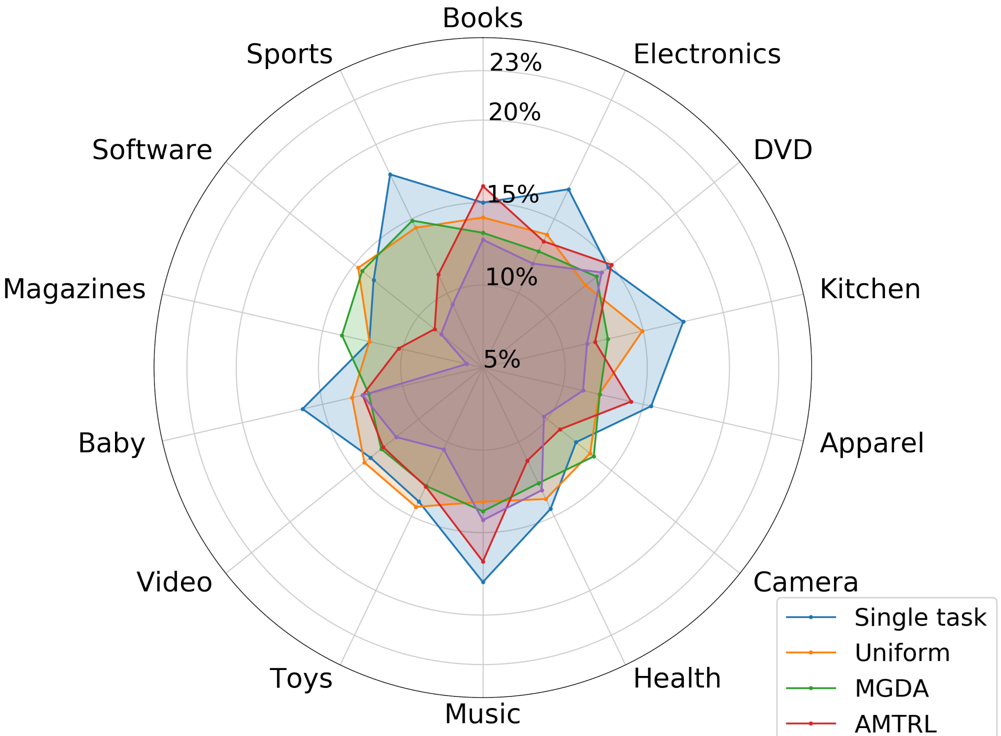
Sentiment Analysis.



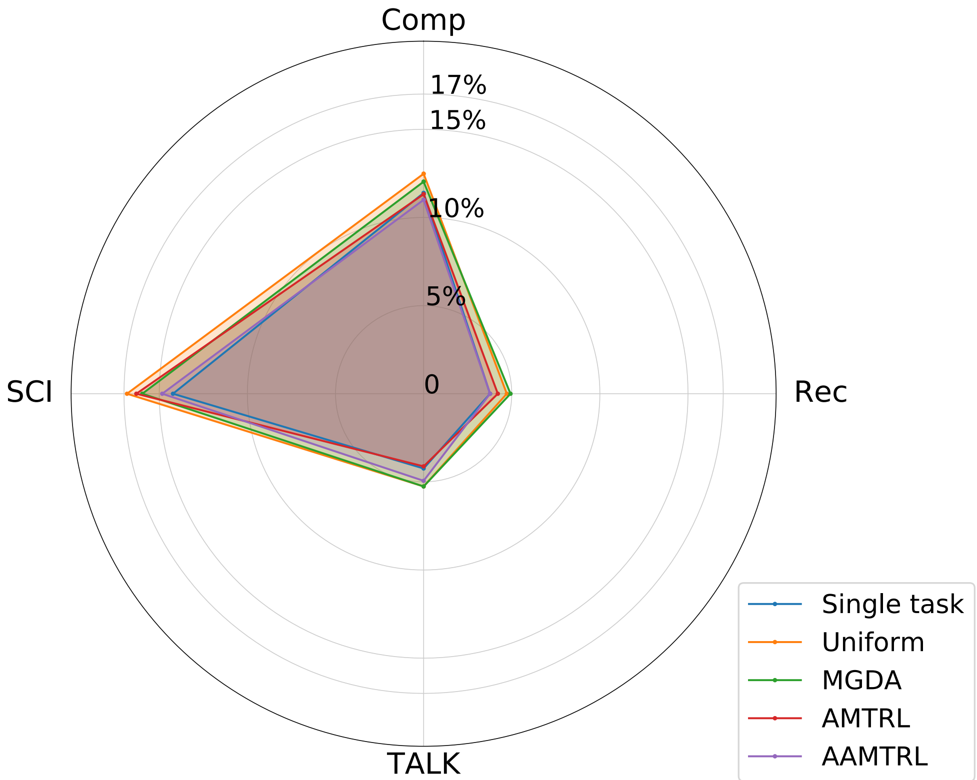
Topic Classification

Experiments - *Classification Accuracy*

Sentiment Analysis and Topic Classification.



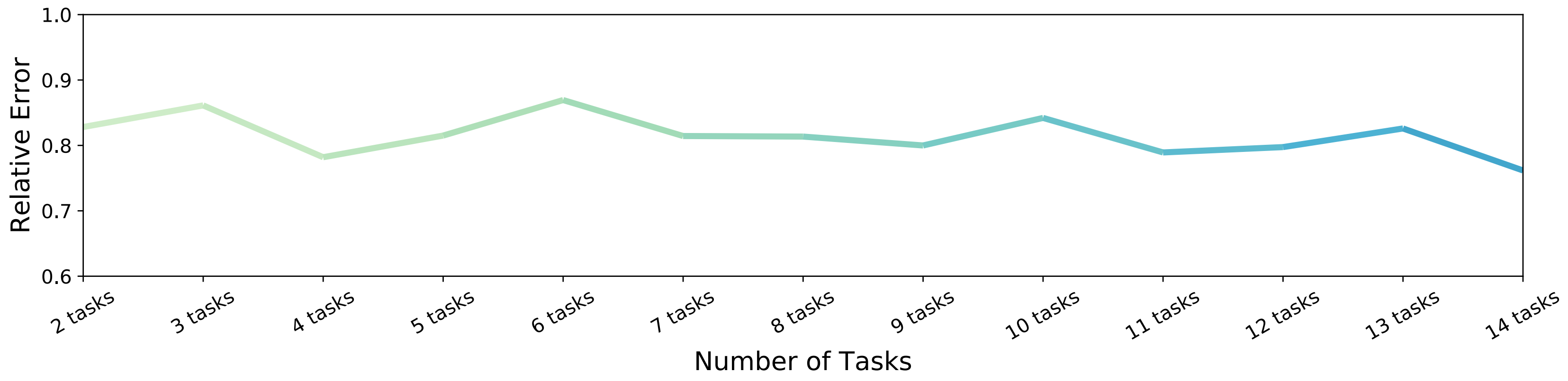
Sentiment Analysis.



Topic Classification

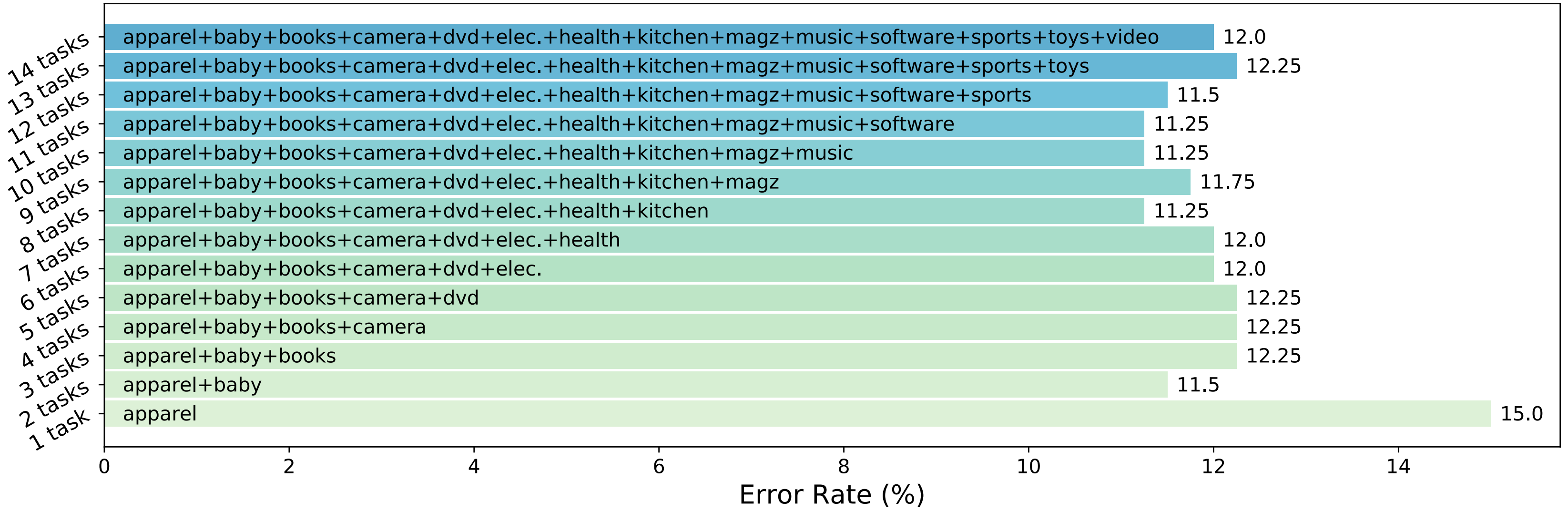
Experiments - *Influence of the Number of Tasks*

Sentiment Analysis.



Relative Error:

$$er_{rel} = \frac{er_{MTL}}{\frac{1}{T}\sum_{1}^{T} er_{STL}^{t}}$$



Error rate for the task 'appeal'.

THANK YOU