Selective Multi-Task Learning by Coupling Common and Private Representations

Jaisiel Madrid-Sánchez, Emilio Parrado-Hernández and Aníbal R. Figueiras-Vidal
Information and Management Processing Group
Dept. of Signal Processing and Communications

Outline

- Selective knowledge transfer among tasks with varying degrees of relatedness
- Joint functional to be optimized is formed by common and private parts
- Task relatedness expressed varying the contributions of these parts
- Two MTL frameworks under the proposed paradigm:
  - Selective Multitask Single Layer Perceptron
  - Selective Multitask Multiple Layer Perceptron
- By product of the MLP case: Implicit internal feature selection.

1st framework

- Hypothesis is a direct mapping from the input space $X$ onto the output space $Y$

$$X \xrightarrow{H} Y$$

- Selective SLP-MTL: Hyperbolic tangent activation function composed of private and common parts

$$h_i(x_c) = \tanh((w_c + w_i)x_c)$$

- Joint cost function to obtain all the hypotheses corresponding to $T$ tasks performing selective transfer

$$\min_{\mathbf{w}_T, \rho} \left\{ \sum_{t=1}^{T} \sum_{i=1}^{M} (h_i(x_c) - p_t)^2 + \lambda \left[ \|w_c\|_2^2 + \sum_{t=1}^{T} \|w_t\|_2^2 \right] \right\}$$

- Selective transfer among tasks controlled with parameter $\rho$
  - Large value of $\rho \Rightarrow$ Single Task Learning
  - Small value of $\rho \Rightarrow$ Same model ($w_t$) for all tasks

2nd framework

- Algorithms that construct compound mappings from the input space to some latent space(s) and from the latent space(s) to the output space

$$X \xrightarrow{R} V \xrightarrow{G} Y$$

- Not only the internal mappings have a further decomposition into shared and private parts; the internal representation spaces can also be divided into a shared and a private subspaces.

- Selective MLP-MTL: hidden layer serves as an internal representation space $V$, with $k$ neurons/features $\{x_{k,t}\}^k_{t=1}$ for the shared subspace and another $k$ for the private one $\{x_{p,t}\}^k_{t=1}$

- Selective Multitask learning improves Single task performance
- The proposed framework improves over non selective multitask approaches

Experiments (synthetic)

- Main task plus secondary task with variable relatedness degree (by flipping the labels)
- 500 training samples and 2000 test samples

- Selective transfer enables positive transfer from secondary to main when relatedness is high
- Selective transfer decouples secondary task when relatedness is low
- Implicit Internal Feature Selection

Experiments (benchmarks)

- Dermatology (6 tasks, 200 training samples, 166 for test) and Monks (3 tasks, 124 tr, 432 tst)
- Coupling parameter learnt through cross validation

<table>
<thead>
<tr>
<th>Method</th>
<th>Dermatology (Error (%)</th>
<th>Monks (Error (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear STL</td>
<td>1.12 ± 0.14</td>
<td>1.25 ± 0.18</td>
</tr>
<tr>
<td>Non-linear STL</td>
<td>1.14 ± 0.09</td>
<td>1.16 ± 0.09</td>
</tr>
<tr>
<td>Linear MTL</td>
<td>1.19 ± 0.20</td>
<td>1.21 ± 0.20</td>
</tr>
<tr>
<td>Non-linear MTL</td>
<td>1.16 ± 0.25</td>
<td>1.62 ± 0.32</td>
</tr>
<tr>
<td>Selective MTL</td>
<td>0.98 ± 0.08</td>
<td>0.82 ± 0.08</td>
</tr>
<tr>
<td>Private MTL</td>
<td>0.97 ± 0.05</td>
<td>0.85 ± 0.05</td>
</tr>
<tr>
<td>Selective MLP</td>
<td>0.98 ± 0.08</td>
<td>0.85 ± 0.05</td>
</tr>
</tbody>
</table>

- Selective Multitask learning improves Single task performance
- The proposed framework improves over non selective multitask approaches

Conclusions and Ongoing Research

- Task relatedness introduced in the joint functional with a learnable coupling parameter. Allows to switch between single task learning and classic multitask learning
- Two frameworks to combine common and private features in the jointly learned hypothesis.
  - Implicit internal feature selection
- Ongoing research:
  - Adapt framework to other learning algorithms (SVM) and settings (Multiview learning)
  - Secondary cost function on the coupling parameter to control the selective transfer.