MetaReg: Towards Domain Generalization using Meta-Regularization

Yogesh Balaji\textsuperscript{1}, Swami Sankaranarayanan\textsuperscript{2} and Rama Chellappa\textsuperscript{1}

\{yogesh, rama\}@umiacs.umd.edu, swamiviv@butterflynetinc.com
\textsuperscript{1}University of Maryland, College Park \textsuperscript{2}Butterfly Network Inc, NY

Goal

\begin{itemize}
\item Shift in data distributions results in performance drop in cross-domain recognition tasks
\item Domain generalization aims at generalizing to novel test domains using variations from multiple source domains
\end{itemize}

Main insights

\begin{itemize}
\item Use of regularization function to improve performance in novel test conditions
\item Propose a parameteric regularization function acting on the weights of a neural network to improve cross-domain generalization
\item Parameters of the regularizer estimated using meta-learning
\item Final model trained by minimizing regularized cross-entropy loss using the learnt regularization function.
\end{itemize}

Algorithm

Parameters of feature network $F$: $\psi_i$, task network $T_i$: $\theta_i$, regularizer $R$: $\phi$

Loss function of domain $i$:

$$L_{(i)}(\psi, \theta) = \mathbb{E}_{(x,y) \sim D_i}[-y \cdot \log(T_i \circ F_{\theta_i}(x))]$$

Base network optimization:

$$\min_{\theta_i, \phi} \sum_{i=1}^{D} L_{(i)}(\psi, \theta)$$

Base network parameters at iteration $k$:

$$[\theta_{(k)}^1, \ldots, \theta_{(k)}^D]$$

Optimizing the regularizer (Updates for generalizing from domain $a$ to $b$):

$$\beta^l \leftarrow \beta^l - \alpha \nabla_{\theta^l} \sum_{i \in \{a, \ldots, l\}} L_{(i)}(\psi, \theta)$$

$$\beta^l = \beta^{l-1} - \alpha \nabla_{\theta^l} \left( L_{(a)}(\psi^a, \theta^a) + R_{(\beta^{l-1})} \right) \quad \forall l \in \{2, \ldots, l\}$$

$$\theta^l_{(k+1)} = \theta^l_{(k)} - \alpha \nabla_{\theta^l} L_{(i)}(\psi^i, \theta^l_{(k)}) \quad (Repeat \ these \ updates \ for \ every \ (a, b) \ pair)$$

Training the final model:

$$\min_{\theta_i, \phi} \sum_{i=1}^{D} L_{(i)}(\psi_i, \theta_i) + R_{(\phi)}$$

Experiments

Recognition accuracies (in %) on PACS dataset using Alexnet:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
Method & Art painting & Cartoon & Photo & Sketch & Average \\
\hline
Base & 76.2 & 83.7 & 72.2 & 69.8 & 75.7 \pm 0.35 \\
\hline
MetaReg (Ours) & \textbf{83.1} & \textbf{87.2} & \textbf{79.2} & \textbf{79.7} & \textbf{81.3} \pm 0.22 \\
\hline
\end{tabular}

Recognition accuracies (in %) on PACS dataset using Resnet:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
Method & Art painting & Cartoon & Photo & Sketch & Average \\
\hline
Base & 79.9 & 77.2 & 72.8 & 70.3 & 75.1 \pm 0.33 \\
\hline
MetaReg (Ours) & \textbf{81.4} & \textbf{87.2} & \textbf{79.2} & \textbf{79.7} & \textbf{82.8} \pm 0.20 \\
\hline
\end{tabular}

Recognition accuracies (in %) on Amazon reviews dataset:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
Method & Basics & DTD & Electronics & Kitchen & Average \\
\hline
Base & 75.5 & 80.7 & 78.5 & 79.0 & 78.3 \pm 0.36 \\
\hline
MetaReg (Ours) & \textbf{76.1} & \textbf{80.6} & \textbf{83.9} & \textbf{83.1} & \textbf{82.7} \pm 0.26 \\
\hline
\end{tabular}

References

[3] Li et al., Deeper, broader and artier domain generalization, ICCV 2017
[4] Li et al., Learning to generalize: Meta-learning for domain generalization, AAAI 2018

Histgram of weights of the task network

The authors acknowledge support of Army Research Office under the Grant No. W911NF-17-1-0304. This is part of the collaboration between US DOD, UK MOD and UK Engineering and Physical Research Council (EPSRC) under the Multidisciplinary University Research Initiative