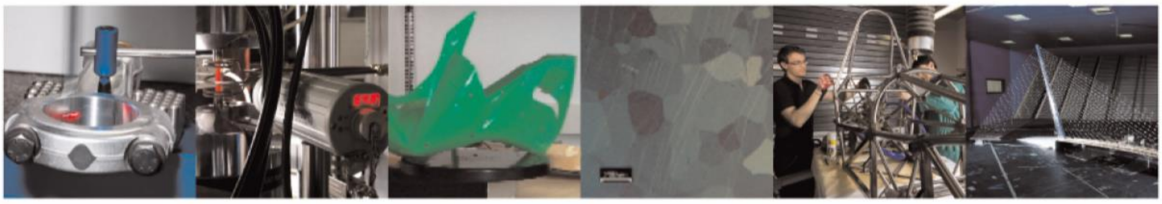

Discussion of “Statistical transfer learning: A review and some extensions to statistical process control”

Tsiamyrtzis, P.

Discussion of the paper: "Statistical transfer learning: a review and some extensions to statistical process control"

Panagiotis Tsiamyrtzis

Department of Statistics, Athens University of Economics and Business

76 Patission Street, Athens 10434, Greece.

pt@aueb.gr

I would like to start by congratulating the authors for the nice presentation on the very interesting topic of Statistical Transfer Learning (STL) and especially for bringing to the fore its relation to the area of Statistical Process Control (SPC). In what follows some further aspects of statistical transfer learning will be presented while others will be discussed further, aiming to offer a more spherical point of view of this area.

1 Aspects of Statistical Transfer Learning

Transfer learning attempts to carry over information among tasks and improve the learning procedure. This sharing of information across tasks (depending on the problem under study) can be done either in parallel or sequentially (see Figure 1).

Parallel STL: When various tasks need to be learned simultaneously. These tasks might be different spatial/temporal cases of the same task or tasks from different domains. So knowledge transfer is multi-directional, like the landslide monitoring example.

Sequential STL: When we have one or more tasks that we have learned and we are interested in transferring this available knowledge in the learning of a new task. So we have unidirectional transfer, like the 3D printing quality control example.

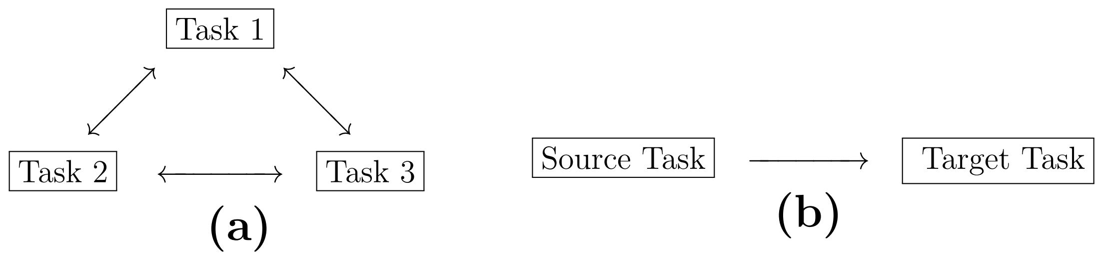

Figure 1: Parallel **(a)** versus sequential **(b)** statistical transfer learning.

The former (also known as multi-task learning in machine learning) has the advantage that we have a bigger set to work with but the learning is done simultaneously. On the other hand the latter allows knowledge attained from past (source) data to be carried over to new (target) data analysis.

2 Negative Transfer Learning

In learning a new (target) task one can try to either do it:

(i) from scratch, i.e. use just the available data and ignore any relative knowledge

(ii) use transfer learning to carry over information from source task(s) and attempt to have improved performance in learning the target task.

Is transfer learning always preferable? The answer is no. The scenario where transfer learning can potentially make things worst is known as negative transfer learning (Torrey and Shavlik, 2009), where attempting to transfer knowledge from source to target task not only does not improve performance but it may actually decrease it. To avoid negative transfer the user needs to be careful, that the tasks are "similar" enough and that the transfer method is "well leveraged". This is quite demanding for an autonomous system and methods that will be able to protect against negative transfer (allowing only "safe" transfer) will most likely reduce the benefit of transfer learning, compared to a method that does "aggressive" transfer learning, which will have excellent performance in similar tasks but will allow negative transfer in dissimilar scenarios.

3 Statistical Methods in Transfer Learning

Transfer learning is not really a new concept in the area of statistics. The Bayesian approach for example, can be seen as a transfer learning mechanism. The idea of a prior distribution along with the hierarchical modeling naturally fit this purpose. As a representative example one can refer to the power priors (Ibrahim and Chen, 2000) which play the role of (sequential) statistical transfer learning mechanism: if D_0 are the source task data we form the prior:

$$\pi\left(\theta|D_0,\alpha_0\right) \propto \left[L\left(\theta|D_0\right)\right]^{\alpha_0}\pi_0(\theta)$$

and then upon observing the target data D_1 we obtain the posterior:

$$\pi\left(\theta|D_1,D_0,\alpha_0\right) \propto L\left(\theta|D_1\right)\left[L\left(\theta|D_0\right)\right]^{\alpha_0}\pi_0(\theta)$$

where the value of $\alpha_0 \in [0,1]$ will determine the effect (reflecting the similarity between the source and the target task) of the source data (D_0) in determining the posterior distribution of the parameter θ, once we use the target data (D_1).

Can the Bayesian approach allow negative transfer? If the source and target tasks have been generated from very different values of parameters then the posterior in the target task can be negatively affected from the prior (that was set from the source task), mainly when we have low volumes of data, as with big data the effect of the prior diminishes. However, prior sensitivity analysis can be helpful to examine whether the prior used, affects (negatively) the posterior or not.

4 Transfer Learning and SPC

The paper demonstrates ways where SPC methods can provide tools in statistical transfer learning. An interesting question though comes if we inverse the above and ask whether SPC methods can benefit from the use of statistical transfer learning. For example in frequentist based control charting (Shewhart charts, CUSUM, EWMA etc.) a standard practice is to employ a phase I/II split, where in phase I we perform learning (calibration) while in phase II we perform testing. So learning stops at the end of phase I. Transfer learning philosophy would suggest to carry over learning in phase II, incorporating the information from new data as they become available. Such a proposal is feasible via a Bayesian SPC scheme (see for example Tsiamyrtzis and Hawkins 2005 & 2010) which can be set as a sequentially updated mechanism allowing the parameter transfer learning, as data become available progressively, providing solutions even when we have small amounts of data and braking free form the usual phase I/II constraint.

References

[1] Ibrahim, J., G. and Chen, M., H. (2000), "Power Prior Distributions for Regression Models", *Statistical Science*, Vol. 15, No. 1, pp. 4660.

[2] Torrey, L. and Shavlik, J. (2009), "Transfer Learning", *Handbook of Research on Machine Learning Applications*, editors: E. Soria, J. Martin, R. Magdalena, M. Martinez and A. Serrano, editors, IGI Global, chapter 11, pp. 242-264.

[3] Tsiamyrtzis, P. and Hawkins, D. M. (2005), "A Bayesian scheme to detect changes in the mean of a short-run process" *Technometrics*, Vol. 47, Issue 4, pp. 446-456.

[4] Tsiamyrtzis, P. and Hawkins, D. M. (2010), "Bayesian startup phase mean monitoring of an autocorrelated process that is subject to random sized jumps", *Technometrics*, Vol. 52, Issue 4, pp. 438-452.