Acta Universitatis Apulensis
ISSN: 1582-5329
http://www.uab.ro/auajournal/

No. 48/2016
pp. 139-146
doi: 10.17114/j.aua.2016.48.12

A REVIEW ON TRANSFER LEARNING IN MEDICAL IMAGING

A. Birlutiu, A. Maier

Abstract. Machine learning techniques are increasingly successful in image-based diagnosis, disease prognosis, and risk assessment. Transfer learning has emerged as a new machine learning framework that uses knowledge from similar learning tasks and domains to increase the performance of a target learning task. Very recently transfer learning has started to be investigated in medical imaging. This paper highlights new research directions related to machine and transfer learning in medical imaging.

2010 *Mathematics Subject Classification*: 30C45, 30C10.

Keywords: medical image analysis, machine learning, transfer learning

1. Introduction

The applications of machine learning to medical engineering and in particular medical image processing have attracted a lot of attention in recent years [12, 13]. Machine learning has opened new ways to extract and interpret the informational content of medical images. Techniques based on machine learning, through a process called training – learning from examples – create models which are capable to identify, classify, and label the image content. These models can be used for computer aided diagnosis, image-based disease prognosis, etc., in general, these models can support the entire clinical imaging workflow, from screening and diagnosis to patient stratification, therapy planning, intervention and follow-up.

There are two major challenges to overcome when applying machine learning methods for medical image analysis. First, the amount of labeled medical data is typically very limited, and a classifier cannot be effectively trained to attain high performance. Second, medical domain knowledge is required to identify representative features in data for performing various tasks, such as disease detection. In the medical domain, training data are generally difficult to acquire because the manual labeling is a complex and time-consuming activity. These limitations are restraining the applications of machine learning in biomedical research and in clinical practice.

An approach to these limitations is to apply a popular machine learning framework, namely transfer learning.

Developing machine learning techniques for medical image analysis has been a hot research topic in the recent years. Only very recently the subfield of transfer learning is getting some attention in the medical image analysis research community [3, 4, 5].

2. Machine Learning

In the recent years, machine learning had quite some successful applications in different areas, such as speech and hand-write recognition, medical diagnosis, computational biology, medical surgery, astronomy, stock market analysis, image processing and recently medical image analysis. Its success can largely be explained by the increasing availability of empirical data and computational power.

Most of the machine learning tasks are supervised in the sense that the learning consists of inferring a model from training data. The training data is formed by a set of examples, each example being a pair made by an input object and a desired output. A supervised learning algorithm analyzes the training data and produces an inferred function, a model which is called a classifier (if the output is discrete) or a regression function (if the output is continuous. The inferred function should predict the correct output value for any valid input object. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.

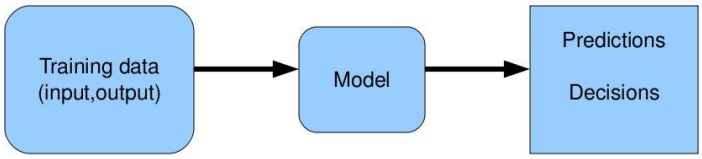

Figure 1: Supervised machine learning. A supervised learning algorithm analyzes the training data and produces a model which is used to make predictions and take decisions.

An important issue that machine learning has to address in many real-world situations, is the limited availability of labeled data used for model training (since this requires either time-consuming and costly laboratory experiments or interactions with a human user).

3. Transfer Learning

Transfer learning has emerged as a new machine learning framework that investigates how to recognize and apply knowledge and skills learned in previous tasks to novel tasks in new domains [1, 2]. It is motivated by human learning, people can intelligently apply knowledge learned previously to solve new problems faster or with better solutions. It is inspired by the research on transfer of learning in psychology, more specifically on the dependency of human learning on prior experience. For example, learning a new language is easier if you speak five languages than when you start your first new language. You use knowledge and experience of the languages that you already know for learning faster and better a new language. The psychological theory of transfer of learning implies the similarity between tasks. In a related way, transfer learning in machine learning assumes similarity between models of different tasks. The transfer learning setting includes a source and a target domain and its corresponding learning tasks. The idea behind transfer learning is that the involved domains share some common latent information (transfer factors/components), which can be exploited by using different techniques as the bridge for knowledge transfer.

Knowledge is extracted from the source domain and transferred to the target domain in order to improve the learning of the target predictive function or make the learning task easier in some way. For example, if very little training data is available in the target domain, and the training data is plentiful in another related domain, transfer learning, if done successfully, can leverage information from the source domain to improve the performance of learning in the target domain.

3.1. Terminology

With respect to terminology, terms such as multi-task learning, domain adaptation, domain generalization, covariate shift, target shit, transferred dimensionality reduction, etc., refer to different data settings and situations but they can all be put under the umbrella of the general transfer learning term.

4. Transfer Learning Approaches

The approaches to transfer learning can be divided into three major categories: 1) instance-transfer methods in which different weights are learned to rank training examples in a source domain for better learning in a target domain; 2) parameter-transfer methods in which the source and target model share some parameters or a prior/hyperprior distribution; and 3) feature-transfer methods which learn a common feature structure that can bridge the source and target domains for knowledge

transfer. In this section we discuss the last two types of approaches and their applicability to medical image analysis. We further discuss the combination of transfer learning with a recent and successful machine learning technique, namely deep learning.

4.1. Parameter-transfer methods

Parameter-transfer methods assume that the source and target model share some parameters which in the Bayesian approach reduces to sharing a prior/hyper-prior distribution. Examples of the Bayesian approach to transfer/multi-task learning are (Bakker and Heskes, 2003) where a mixture of Gaussians is used for the top of the hierarchy. This leads to clustering the tasks, one cluster for each Gaussian in the mixture. In [11] a hierarchical Gaussian Process [22] is derived with a normal-inverse Wishart distribution used at the top of the hierarchy. The normal-inverse Wishart distribution $\mathcal{IW}(\mathbf{\Sigma}|\tau, \mathbf{\Sigma}_0)$ is specified by means of the scale matrix $\mathbf{\Sigma}_0$ with precision τ and mean $\boldsymbol{\mu}_0$ with precision π, and it is chosen since it is the conjugate prior for the multivariate distribution which is used for model parameters. In this way, many of the intractable computations can be performed analytically. The hierarchical prior was used to sample model parameters in order to enforce a similar structure for the utility function of each individual subject.

Let us assume we have to model and learn multiple dependent functions. The functions we need to learn share something in common: for example, this can be the mean of those functions and the local smoothness. Associated with each function we have some observations. Furthermore each function f depends on some parameters theta. We can implement multi-task learning using hierarchical modeling in which the individual parameters θs depend on some common parameter α. We used hierarchical modeling to derive a method for gathering data from other users in a prior information for a new user. This approach has also been shown to work in speech recognition [19, 20].

The Bayesian approach to transfer learning assumes the parameters of individual models to be drawn from the same prior distribution.

4.2. Feature-transfer methods

Feature-transfer methods address the mismatch between source and target data distributions. Source knowledge is used for learning a good feature representation in the target space.

Transfer Component Analysis (TCA) [21] is a feature-transfer technique that can deal with non-linearities and complex changes in the data. TCA learns a shared subspace by minimizing the dissimilarities across domains, while maximally preserving

the data variance.

In [16] Multi-TCA, an extension of TCA to multiple domains, and Multi-SSTCA, an extension of TCA for semi-supervised learning were proposed. The basic idea of domain adaption techniques is to make the source and target data distributions as similar as possible and in the same time preserve properties of the original data. This idea is implemented by learning a shared subspace between the source and target data. Standard machine learning methods, such as support vector machine, random forest, penalized least squares regression, decision trees, ensemble learning, etc., can be used in this subspace to train classifiers or regression models across domains. Let X_S and X_T be the data from source and target domains respectively, and let $P(X_S)$ and $P(X_S)$ be the source and target data probability distributions. TCA assumes that $P(X_S) \neq P(X_T)$. This assumption holds in many medical contexts when data comes from different patients or the medical image data is obtained with different modalities. The basic idea of TCA and related domain adaptation techniques is to assume a transformation $\phi : \mathcal{X} \to \mathcal{H}$ induced by a universal kernel, where $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS) (Hofmann, et al., 2008) such that $P(\phi(X_S)) \approx P(\phi(X_T))$. The mapping ϕ is determined such that: 1) the distance between the marginal distributions $P(\phi(X_S))$ and $P(\phi(X_T)$ is small; and 2) the variance of the original data is preserved (this constraint is similar to Principal Component Analysis). Instead of finding ϕ explicitly, the corresponding kernel matrix [17] associated with ϕ is learned and all the computations are performed using kernels.

In order to make the distributions of data from source and target domains close to each other, a distance measure between distributions is needed. Many criteria, such as the Kullback-Leibler (KL) divergence, can be used. However, many of these criteria are parametric, an intermediate density estimate is usually required, and then matching the real data distribution to the parametric. To avoid such a non-trivial task, a non-parametric distance estimate between distributions is more desirable. Maximum Mean Discrepancy (Gretton, Borgwardt et al, 2007) is a criterion for comparing distributions based on a RKHS which combines very well with the idea of TCA of mapping the data in a RKHS space. The distance between source and target data distributions is estimated by the distance between the means of the two samples mapped into a RKHS.

5. Deep learning

Deep learning [18] is a composite model of neural networks which is recently very successful and is shown to achieve substantial improvements in classifying images, audio, and speech data. An autoencoder is an artificial neural network used to learn

a compressed, distributed representation (encoding) for a set of data. A stacked autoencoder is a neural network consisting of multiple layers in which the outputs of each layer is wired to the inputs of the successive layer. All deep learning models require a substantial amount of training instances to avoid the problem of overfitting.

Deep learning has been proven to be very successful in classifying images, audio, and speech data and is getting attention in medical image analysis [15, 6].

References

[1] A Argyriou, T Evgeniou, M Pontil, *Convex multi-task feature learning*, Machine Learning 73 (3), 243-272, 2008.

[2] S.J. Pan, Q. Yang, *A survey on transfer learning*. IEEE Trans. Knowle. Data Eng. 2010;vol. 22, pp. 1345-1359.

[3] T. Heimann, P. Mountney, M. John, R. Ionasec, *Real-time ultrasound transducer localization in uoroscopy images by transfer learning from synthetic training data*. Medical Image Analysis. 18(8), 1320-1328.

[4] A. van Opbroek, M.A. Ikram, M.W. Vernooij, M. de Bruijne, *Transfer learning improvehys supervised image segmentation across imaging protocols*. IEEE Trans. Med. Imag 34(5): 1018-1030.

[5] Y. Zheng, *Cross-modality medical image detection and segmentation by transfer learning of shape priors*. Proc. IEEE Intl Sym. Biomedical Imaging, 2015.

[6] F. Ghesu, B. Georgescu, Y. Zheng, J. Hornegger, D. Comaniciu, *Marginal Space Deep Learning: Efficient Architecture for Detection in Volumetric Image Data*. 18th International Conference on Medical Image Computing and Computer Assisted Intervention.

[7] K. Muller, G. Lauritsch et al., *Catheter artifact reduction (CAR) in dynamic cardiac chamber imaging with interventional C-arm CT*. Proceedings of the third international conference on image formation in x-ray computed tomography. pp. 418-421, 2014.

[8] J. Almazan, A. Gordo, A. Fornes, E. Valveny, *Word spotting and recognition with embedded attributes*. Pattern Analysis and Machine Intelligence, IEEE Trans on vol. 36, issue 12, pp. 2552 - 2566, 2014.

[9] A. Birlutiu, F. d'Alche-Buc, T. Heskes, *A Bayesian Framework for Combining Protein and Network Topology Information for Predicting Protein-Protein Interactions.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol.12, pp: 538 - 550, 2015.

[10] B. Bakker, T. Heskes *Task clustering and gating for Bayesian multitask learning.* Journal of Machine Learning Research 4:83-99, 2003.

[11] A. Birlutiu, P. Groot, T. Heskes, *Multi-task preference learning with an application to hearing aid personalization*, Neurocomputing, 73 (7), 1177-1185, ISSN: 0925-2312, 2010.

[12] de Bruijne M., *Machine learning approaches in medical image analysis: From detection to diagnosis* Medical Image Analysis Volume 33, October 2016, Pages 9497.

[13] T. Geimer, A. Birlutiu, M. Unberath, O. Taubmann, C. Bert, A. Maier *A Kernel Ridge Regression Model for Respiratory Motion Estimation in Radiotherapy.* Bildverarbeitung fr die Medizin 2017 - Algorithmen Systeme Anwendungen (Workshop Bildverarbeitung fr die Medizin 2017), Heidelberg, 12.-14.03.2017 (accepted).

[14] A. Birlutiu, P. Groot, T. Heskes, *Efficiently learning the preferences of people.* Machine Learning Journal, 90 (1), pp.1-28, Springer, ISSN: 0885-6125, 2013.

[15] H. Greenspan, B. van Ginneken, R. M. Summers, *Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique*, IEEE Transactions on Medical Imaging 35 (5) (2016) 11531159

[16] T. Grubinger, A. Birlutiu, H. Schoner, T. Natschlager, T. Heskes *Domain generalization based on transfer component analysis.* 13th International Work-Conference on Artificial Neural Networks, IWANN Proceedings, Part I. Series Volume: 9094, pp. 325-334, 2015.

[17] T. Hofmann, B. Scholkopf, A. Smola, *Kernel methods in machine learning*, Ann. Statist. Volume 36, Number 3 (2008), 1171-1220.

[18] LeCun Y., Bengio Y., Hinton G.E. *Deep learning.* Nature, Vol. 521, pp 436-444, 2005.

[19] Maier A., *Speech of Children with Cleft Lip and Palate: Automatic Assessment*, ISBN 978-3-8325-2144-8, Publisher: Logos, vol. 29, 2009.

[20] Maier A., Haderlein T., Nöth E., *Environmental Adaptation with a Small Data Set of the Target Domain*, Lecture Notes in Artificial Intelligence, Text, Speech and Dialogue, editor: Petr Sojka and Ivan Kopecek and Karel Pala, ISSN: 0302-9743, vol. 1, pp. 431–437, 2006.

[21] S. Pan, I. Tsang, J.T. Kwok, Q. Yang, *Domain adaptation via transfer component analysis.* IEEE Transactions on Neural Networks, 199-210, 2011.

[22] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.

Adriana Birlutiu
Department of Mathematics and Informatics,

"1 Decembrie 1918" University of Alba Iulia,
5, Gabriel Bethlen street
510009 Alba Iulia, Romania
email: *adriana.birlutiu@uab.ro*

Andreas Maier
Computer Science Department
Pattern Recognition Lab
Friedrich-Alexander-Universitt Erlangen-Nurnberg
Martensstrasse 3,
91058 Erlangen, Germany
email: *andreas.maier@fau.de*