Pattern Recognition 73 (2018) 65–75

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

On automated source selection for transfer learning in convolutional neural networks

Muhammad Jamal Afridi[a,*], Arun Ross[a], Erik M. Shapiro[b]

[a] Department of Computer Science and Engineering, Michigan State University, USA
[b] Department of Radiology, Michigan State University, USA

ARTICLE INFO

Article history:
Received 13 January 2017
Revised 23 April 2017
Accepted 17 July 2017
Available online 26 July 2017

Keywords:
Transfer learning
CNN selection
Deep learning

ABSTRACT

Transfer learning, or inductive transfer, refers to the transfer of knowledge from a source task to a target task. In the context of convolutional neural networks (CNNs), transfer learning can be implemented by transplanting the learned feature layers from one CNN (derived from the source task) to initialize another (for the target task). Previous research has shown that the choice of the source CNN impacts the performance of the target task. In the current literature, there is no principled way for selecting a source CNN for a given target task despite the increasing availability of pre-trained source CNNs. In this paper we investigate the possibility of automatically ranking source CNNs prior to utilizing them for a target task. In particular, we present an information theoretic framework to understand the source-target relationship and use this as a basis to derive an approach to automatically rank source CNNs in an efficient, zero-shot manner. The practical utility of the approach is thoroughly evaluated using the Places-MIT dataset, MNIST dataset and a real-world MRI database. Experimental results demonstrate the efficacy of the proposed ranking method for transfer learning.

1. Introduction

1.1. Background and motivation

Deep learning methods, specifically those based on convolutional neural networks (CNNs), have demonstrated tremendous success in a variety of applications ranging from object recognition to autonomous driving [1,2]. One key reason behind this unprecedented success is the availability of large application-specific, annotated datasets. However, in many practical applications, especially those related to medical imaging and radiology, obtaining a large annotated (e.g., labeled) dataset can be challenging. In many cases, annotation can only be performed by qualified field experts and so crowd sourcing methods, such as Amazons Mechanical Turk [3], cannot be used for annotating data. These limitations can often preclude the use of CNNs in such applications.

In order to address the problem of limited training data, the concept of *transfer learning* can be used. In transfer learning, knowledge learned for performing one task is used for learning a different task. The idea of transfer learning is not new. For example, the NIPS'95 workshop on *Learning to Learn* highlighted the importance of pursuing research in transfer learning. A number of research studies have been published in the past investigating different aspects of transfer learning. Some of these studies have been summarized in Table. 1. Based on *what* is transferred, these approaches can be mainly categorized as (1) instance-based transfer learning, where the *labeled data* in the source task is re-weighted to be utilized for the target task [4–7], (2) feature-based transfer learning, where the *features* of the source task are transformed to closely match those of the target task, or a common latent feature space is discovered [8–10], (3) parameter-based transfer learning, where the goal is to discover *shared parameters* across tasks [11,12] and (4) relational knowledge-based transfer learning, which is a comparatively less explored area in this context, and where the goal is to transfer the *relationship* among data from a source task to a target task [13].

In case of CNNs, transfer learning typically entails the transfer of information from a selected source concept (source CNN, *learned for a source task*) to learn the target concept (target CNN, *learned for a target task*). Recent studies detail how transfer learning can be performed via CNNs by transplanting the learned feature layers from one CNN to initialize another [22]. Due to its significant

* Corresponding author.
E-mail addresses: afridimu@msu.edu, jamal.afridi2@gmail.com (M.J. Afridi), rossarun@msu.edu (A. Ross), shapir86@msu.edu (E.M. Shapiro).

http://dx.doi.org/10.1016/j.patcog.2017.07.019
0031-3203/

Table 1
A brief overview of transfer learning research.

Paper	Focus of research	Transfer
Dai et al. [4]	Transfer learning via boosting algorithm	Instance based
Jiang et al. [5]	Source instance weighting for domain adaptation	Instance based
Liao et al. [6]	Utilizing auxiliary data for target labeling	Instance based
Wu and Dietterich [7]	Integrating source task data in SVM learning framework	Instance based
Pan et al. [8]	Transfer learning via dimensionality reduction	Feature based
Pan et al. [9]	Domain adaptation using efficient feature transformation	Feature based
Blitzer et al. [14]	Extracting features to reduce difference between domains	Feature based
Dai et al. [15]	Labeling target task data using unlabeled source task data	Feature based
Duame [16]	Domain adaptation using feature augmentation	Feature based
Xing et al. [17]	Correcting the predicted labels of shift-unaware classifier	Feature based
Pan et al. [18]	Spectral feature alignment for transfer learning	Feature based
Raina et al. [10]	Learning high-level features for transfer learning	Feature based
Gong et al. [19]	Reducing domain difference in a low dimensional feature space	Feature based
Tommasi et al. [11]	Transferring SVM hyperplane information	Parameter based
Yao and Doretto [12]	Transferring internal learner parameter information	Parameter based
Mihalkova et al. [13]	Markov logic networks for transferring relational knowledge	Relational knowledge
Long et al. [20]	Joint domain adaptation	Feature based
Ammar et al. [21]	Automated source selection in reinforcement learning using RBMs	Parameter based

impact on improving the performance of the target task, transfer learning is becoming a critical tool in many applications [23,24]. Usually this process is referred to as *fine-tuning* to indicate that the transplanted feature layers of a source CNN are merely refined using the target data. It is necessary to note that for such a transfer, the source *data* is not needed; only the source *concept* as embodied by the source CNN is required. This allows researchers to freely share and reuse previously learned CNN models.[1] Attempts to convert CNN models from one programming platform to another[2] has also facilitated the reusability of CNNs. Given these developments, it has become necessary to investigate how CNN models learned on various source tasks can be effectively used when learning a target task that has very limited training data.

Given a selected source task or a source CNN, recent studies show a number of useful ways to transfer and exploit its information for maximizing the performance gain on the target task [22,25–28]. Previous research has clearly demonstrated that the *choice* of the source CNN has an impact on the performance of the target task [22]. Some sources[3] may also result in a phenomenon called *negative transfer* where the performance on the target task is degraded as a result of transfer learning. However, a principled reason for such a degradation has not been clearly determined. Further, in CNN-based transfer learning, the source is *manually* chosen (e.g., [23,24]). Several different approaches have been suggested to manually select a source for transfer learning. In [23], Agrawal et al. demonstrate that source data obtained from a moving vehicle [29] can be effective for transfer learning, thereby highlighting the importance of motion-based data. In [22], Yosinki et al. argue that source tasks that appear to be semantically relevant to the target task would result in better performance. A large number of studies, however, show that semantic relevance between source and target tasks is not always necessary; performance improvement has been observed even when the source and target tasks are superficially not related [23,30].

Manual selection has three major drawbacks: it is *subjective*, where multiple experts may choose a different source for the same target task; *unreliable*, where there is no guarantee that the chosen source will result in better performance than others; and *laborious*, where an expert has to manually analyze a very large number

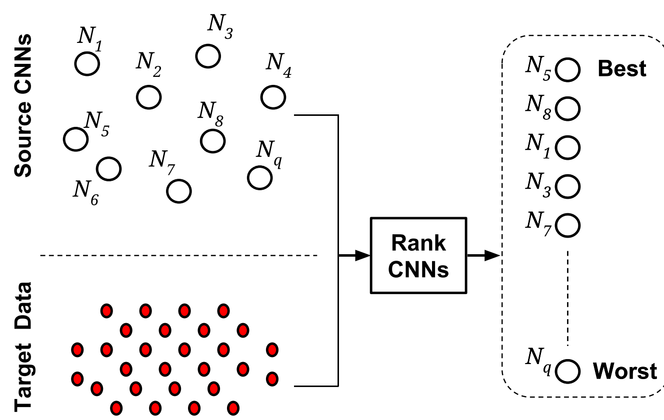

Fig. 1. Given a large number of pre-trained source CNNs, the proposed approach ranks them in the order in which they are likely to impact the performance of a given target task. The source task data is *not* used in this determination.

of potential sources tasks. Currently, there is no principled way to automatically select the best source CNN for a given target task.

1.2. Technical goal

The key technical goal of this study, therefore, is to investigate the possibility of automating source CNN selection (see Fig. 1). By choosing the best source CNN for a given target task, we anticipate that high performance can be achieved despite tuning with very limited target data. Since, this is the first study attempting to automate source CNN selection, we first present the following three ideal requirements of such a ranking measure:

Scalable: It should only utilize source CNNs. It should not require us to additionally store and maintain the source *data* of each source task.

Efficient: Unlike a standard learning based problem where an objective function is defined and optimized using a training dataset, the ideal ranking approach should perform a *zero-shot* ranking of CNNs, i.e., the ranking approach should not utilize a learning phase that is based on source CNN characteristics.

Reliable: Ideally, the ranking measure should not be based on heuristics, especially those simply based on the notion of perceived similarity or difference between the tasks. The ranking measure should be theoretically derived using well-understood principles. The proposed measure shall also demonstrate its efficacy for a practical application.

[1] https://github.com/BVLC/caffe/wiki/Model-Zoo
[2] https://github.com/facebook/fb-caffe-exts
[3] Note that in this paper, "source" will be used as a general term referring to both source task and source CNN.