Pattern Recognition Letters 72 (2016) 82–90

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Transfer metric learning for action similarity using high-level semantics☆

Ziad Al-Halah*, Lukas Rybok, Rainer Stiefelhagen

*Karlsruhe Institute of Technology, Institute for Anthropomatics and Robotics, Karlsruhe 76131, Germany*

ARTICLE INFO

*Article history:*
Available online 17 July 2015

*Keywords:*
Transfer learning
Hierarchical representation
Attributes
Metric learning
Action similarity

ABSTRACT

The goal of transfer learning is to exploit previous experiences and knowledge in order to improve learning in a novel domain. This is especially beneficial for the challenging task of learning classifiers that generalize well when only few training examples are available. In such a case, knowledge transfer methods can help to compensate for the lack of data. The performance and robustness against negative transfer of these approaches is influenced by the interdependence between knowledge representation and transfer type. However, this important point is usually neglected in the literature; instead the focus lies on either of the two aspects. In contrast, we study in this work the effect of various high-level semantic knowledge representations on different transfer types in a novel generic transfer metric learning framework. Furthermore, we introduce a hierarchical knowledge representation model based on the embedded structure in the semantic attribute space. The evaluation of the framework on challenging transfer settings in the context of action similarity demonstrates the effectiveness of our approach compared to state-of-the-art.



## 1. Introduction

Instead of learning new concepts in isolation, humans have the ability to consider connections to previously obtained skills and experiences, which makes our learning process extremely efficient [33]. In psychology, this skill is known as *knowledge transfer* or *transfer learning* [41]. It gives us humans the advantage of learning new concepts faster and with a high initial performance when using only a few trials or examples [39]. In contrast, most machine learning algorithms require a large number of training examples, since training only relies on domain specific data, instead of incorporating prior knowledge [12]. However, in cases when training data is scarce or not available, such methods cannot be applied or are unable to extract a useful model, and thus fail to generalize well. Therefore, there is a growing interest in the Machine Learning Community to mimic this human ability. A typical task that benefits from knowledge transfer is one- and zero-shot learning [5,11,21].

Another problem of many machine learning models is their assumption that training samples are drawn according to the same probability distribution as the unseen test samples [40]. Nevertheless, this hypothesis does not always hold in practical problems, resulting in a reduction of generalization properties. For instance, consider you have built a robust classifier to distinguish between different sports actions and would like to use the same system on more general videos found on YouTube. Usually, this would require an expensive data collection and annotation process. However, using transfer learning methods, it is possible to re-use an established model to save a significant amount of labeling effort [31].

According to [31], transfer learning research tries to solve one or more of the following three problems:

1. *"What to transfer?"* asks what type of knowledge representation is most suitable to be transferred across domains. Hence, an important feature of the transferred knowledge is its ability to encode information that is usable and shareable between tasks.
2. *"How to transfer?"* asks how the transferred knowledge from the source domain can be incorporated in the learning of the target task.
3. *"When to transfer?"* asks when transfer learning is beneficial, since knowledge transfer can sometimes decrease the effectiveness of learning in the target domain (negative transfer). This can for instance happen, when the source and target tasks are very different.

The focus of our work lies on the type of information to transfer across domains, hence on answering the question: *What to transfer?* and consequently of its effect on different types of transfer methods. There are three common approaches in that direction:

1. *Feature representation transfer*, where a knowledge representation model is learned or adopted for the target domain, based on relevant information in the source domain [24,30].

☆ This paper has been recommended for acceptance by Anders Heyden.

* Corresponding author. Tel.: +49 72160844735; fax: +49 72160845939.
*E-mail address:* ziad.al-halah@kit.edu (Z. Al-Halah).

http://dx.doi.org/10.1016/j.patrec.2015.07.005

2. *Parameter transfer*, where models (or parameters) are learned in the source domain and then used to regularize or to be included as a prior in the model learning of the target task [28].
3. *Instance transfer*, where all or some of the samples in the source domain are re-used in the learning of the target task in order to overcome the low number of target training samples [20].

Unlike previous works, which analyze these transfer types separately (e.g. [20,21,44]), we believe that they should be considered jointly. The choice of the feature representation and how the knowledge is modeled will eventually influence the efficiency of all three approaches: the representation-, instance- and parameter transfer. For example, when using color distributions learned on sea animals as a low-level representation, this most likely will generalize poorly to a domain of bird categories and result in a bad performance. However, learning the meta-relations between the categories or the visual semantic attributes (e.g. *has-head*, *is-round* and *has-stripes*) would result in constructing a knowledge space that can be easily shared between various domains. Such high-level semantics are less likely to be influenced by the low-level feature distribution, and consequently form an adequate knowledge representation to be transferred across domains. In transfer metric learning literature, this observation is usually ignored and instead the focus lies on parameter transfer while only using a low-level knowledge representation [43,44].

Another common assumption in the transfer learning literature is that the source data set is much more diverse and complex than the target set and thus the experimental evaluation protocol is designed accordingly (e.g. [11,21,35]). However, collecting and annotating new data is an expensive effort. While we might create data sets of hundreds of action categories there is still tens of thousands of “unseen” classes (i.e. with no training examples). Hence, it seems that it is more likely that we will have a small and simple source domain against a large and diverse target domain. Moreover, the usual case in most research fields is to first focus on solving simple problems before moving on to more complex ones. For instance, the action recognition community started with the task of classifying simple actions in controlled environments (e.g. [38]) and then slowly moved to the complex Action Similarity Labeling (ASLAN) Challenge proposed by Kliper-Gross et al. [19], and beyond. Thus, it would be beneficial if each time we switch to a more challenging task, all previously collected data and experience could be successfully used to improve task performance in the new complex domain. Therefore, we address in our work an evaluation setup where the number and complexity of categories in the source domain is much lower than in the target domain. Such a setup imposes a greater challenge to transfer learning approaches.

In conclusion, the contribution of our work is as follows:

- We show the benefits of using high-level semantics for transfer metric learning.
- We propose a novel hierarchical knowledge representation that encodes the embedded semantic structure of category similarities in the attribute space, and show its superior performance to other semantic models.
- We introduce a novel generic framework for transfer metric learning that improves the transfer performance and reduces the negative transfer effect.
- We suggest a realistic and challenging evaluation protocol for transfer learning, where the target domain is much more diverse and complex than the source domain.

This work is an extended version of [2]. The main additional contribution is an extended evaluation, and discussion of the results. In the added experiments, the emphasis lies on the analysis of how the knowledge complexity of the source sets affects the transfer process.

## 2. Related work

Transfer learning has attracted a lot of attention in the last years, and several approaches were proposed in various fields. Since it is out of scope of this work to summarize all past research efforts, we refer the interested readers to the comprehensive surveys by Pan and Yang [31], and Cook et al. [8], and focus on the most related sub-fields.

### 2.1. Transfer metric learning

While standard supervised and semi-supervised metric learning are widely popular (cf. the survey by Bellet et al. [6]), to the best of our knowledge only two works exist that analyze the application of metric learning to the problem of knowledge transfer. Zha et al. [43] propose to integrate multiple source metrics into a regularized metric learning framework and make use of log-determinant regularization to minimize the divergence between the source metrics and the target metric. A drawback of this approach is that it can only represent positive and zero task correlation, but not negative task correlations. Therefore, Zhang and Yeung [44] proposed a unified framework, called Transfer Metric Learning (TML), that models all three task correlations, while also guaranteeing to find a globally optimal solution. TML is formulated as a special case of multi-task learning, where several independent source tasks and one target task are given, and the relations between the sources and the target are jointly modeled when learning the target metric matrix. Compared to the work of Zha et al. [43], TML showed a superior performance when the training data is scarce. Nonetheless, unlike our work, both approaches use parameter transfer based solely on a low-level feature representation. To the best of our knowledge, the use of high-level semantics and the analysis of the impact of different knowledge representations on the different transfer types have not been addressed before in the context of transfer metric learning.

### 2.2. Knowledge representation transfer

Most of the previous work tackles the distribution differences between the source and target domain as a domain adaptation problem of the low-level features [16,30] or by learning a robust and transferable sparse representation [25]. In contrast to this line of research, we study in this work the robustness of high-level semantic representations in challenging transfer settings. Unlike the common case of domain adaption, transferring high-level knowledge representation does not require the availability of target data at time of representation learning which facilitates and generalizes the transfer process. Moreover, as we will show later in the evaluation, high-level semantics exhibit better performance when transferred across data sets compared to low-level features.

Among the various knowledge models that were introduced recently in the literature, semantic attributes have gained an increasing amount of attention. They describe the visual appearance of an entity and represent an intermediate semantic layer between the low-level features and class categories. Attributes were successfully used in transfer learning applications, like zero-shot recognition of objects, and actions [11,21,23]. Another approach to represent an instance of an unseen class is by its similarity to known categories. This has been applied by Bart and Ullman [5] to one-shot object recognition resulting in a significant improvement in classification performance compared to low-level features.

On the other hand, compared to previous representations, hierarchies proved to be effective due to their ability to capture information at different resolution levels. In fact, there is evidence from neuroscience, that information in the visual cortex is structured hierarchically, e.g. for the high-level tasks of recognizing objects [34] or actions [15]. The structure is usually either defined manually [45], derived from external lexical resources like WordNet [3,35], or based