Neurocomputing 000 (2018) 1–12

Contents lists available at ScienceDirect

Neurocomputing

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

Expectation maximization transfer learning and its application for bionic hand prostheses

Benjamin Paaßen[a,*], Alexander Schulz[a], Janne Hahne[b], Barbara Hammer[a]

[a] Machine Learning Group, Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Inspiration 1, Bielefeld 33619, Germany
[b] Neurorehabilitation Systems Research Group, Department of Trauma Surgery, Orthopedic Surgery and Hand Surgery, University Medical Center Göttingen, Von-Siebold-Str. 3, Göttingen 37075, Germany

ARTICLE INFO

Article history:
Received 9 July 2017
Revised 2 November 2017
Accepted 22 November 2017
Available online xxx

Keywords:
Transfer learning
Gaussian mixture models
Learning vector quantization
Expectation maximization
Bionic prostheses

ABSTRACT

Machine learning models in practical settings are typically confronted with changes to the distribution of the incoming data. Such changes can severely affect the model performance, leading for example to misclassifications of data. This is particularly apparent in the domain of bionic hand prostheses, where machine learning models promise faster and more intuitive user interfaces, but are hindered by their lack of robustness to everyday disturbances, such as electrode shifts. One way to address changes in the data distribution is transfer learning, that is, to transfer the disturbed data to a space where the original model is applicable again. In this contribution, we propose a novel expectation maximization algorithm to learn linear transformations that maximize the likelihood of disturbed data according to the undisturbed model. We also show that this approach generalizes to discriminative models, in particular learning vector quantization models. In our evaluation on data from the bionic prostheses domain we demonstrate that our approach can learn a transformation which improves classification accuracy significantly and outperforms all tested baselines, if few data or few classes are available in the target domain.

1. Introduction

Classical machine learning theory relies on the assumption that training and test data stem from the same underlying distribution; an assumption, that is oftentimes violated in practical applications [1]. The reasons for such violations are multifold. The training data may be selected in a biased way and not represent the "true" distribution properly [1], disturbances may lead to changes in the data over time [2], or one may try to transfer an existing model to a new domain [3]. If such violations occur, the model may not accurately describe the data anymore, leading to errors, e.g. in classification.

This is particularly apparent in the domain of bionic hand prostheses. By now, research prototypes of such prostheses feature up to 20 active degrees of freedom (DoF), promising to restore precise and differentiated hand functions [4]. However, controlling this many degrees of freedom requires a user interface which reacts rapidly and is intuitive to the user. A popular approach to achieve such a user interface is to let users execute the desired motion with their phantom hand, which is still represented in the brain, and infer the desired motion via classification of the residual muscle signals in the forearm, such that the desired motion can then be executed by a bionic hand prosthesis in real-time (time delay below 200 ms) [5]. More precisely, if a user executes a motion with her phantom hand, the corresponding neurons in the brain are activated and propagate the motor command to the arm, where the residual muscles responsible for the hand motion are activated. This activity can be recorded via a grid of electromyographic (EMG) electrodes placed on the skin around the amputee's forearm (see Fig. 1, top left). The EMG signal contains information about the firing pattern of the motor neurons, which in turn codes the intended hand motion. Therefore, one can classify the EMG signal with respect to the intended hand motion and use the classification result to control a prosthesis with little time delay in an intuitive way [5].

Unfortunately, such user interfaces are seriously challenged by changes in the input data distribution due to disturbances to the EMG signal, for example by electrode shifts, posture changes, sweat, fatigue, etc. [5,6]. As an example, consider Fig. 1, which

* Corresponding author.
E-mail addresses: bpaassen@techfak.uni-bielefeld.de (B. Paaßen), aschulz@techfak.uni-bielefeld.de (A. Schulz), janne.hahne@bccn.uni-goettingen.de (J. Hahne), bhammer@techfak.uni-bielefeld.de (B. Hammer).

https://doi.org/10.1016/j.neucom.2017.11.072
0925-2312/

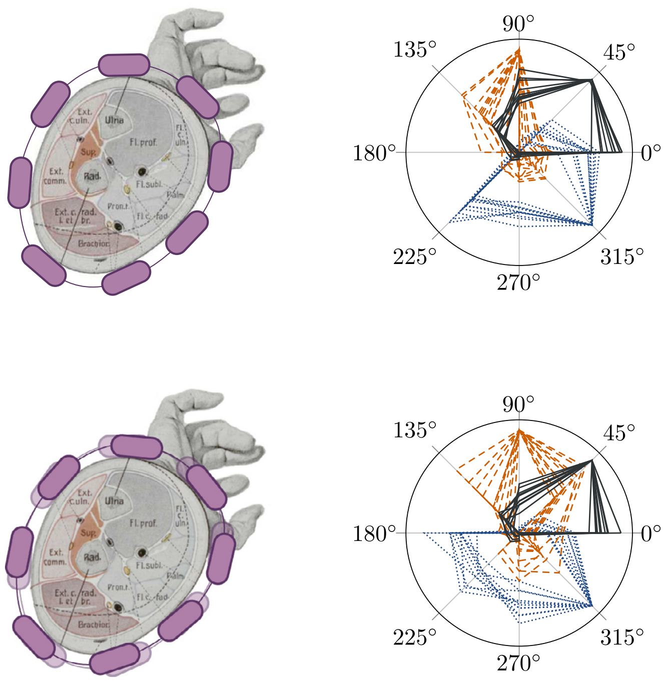

Fig. 1. An illustration of electrode shifts in electromyographic (EMG) data. Top left: a grid of eight EMG electrodes placed around the forearm of a user. Cross section of the arm taken from the 1921 German edition of "Anatomie des Menschen", which is in the public domain. Top right: example EMG signals from an eight-electrode EMG recording for two different hand motions (dashed and dotted lines) as well as resting (solid lines). Bottom left: the electrode grid is shifted around the forearm (electrode shift). Bottom right: another set of EMG signals from a shifted eight-electrode EMG recording for two different hand motions (dashed and dotted lines) as well as resting (solid lines). Due to the shifted signal, a model trained on the source data (top right) may misclassify shifted data (bottom right).

illustrates the effect of an electrode shift around the forearm, leading to different EMG sensor data, which in turn may cause an erroneous classification decision.

Changes between training and test distribution have been addressed by different theoretical frameworks. Shimodaira has introduced the notion of *covariate shift* describing the case of a change in the prior distribution $p(\vec{x})$ while the conditional distribution of the label $P(y|\vec{x})$ remains unchanged [7]. A slightly different angle is taken by *sample selection bias correction theory* which assumes that a *true* underlying distribution $P(y, \vec{x})$ exists from which some pairs are not available in the training data, thereby biasing the resulting machine learning model [1]. In contrast, the theory of *concept drift* models the prior distribution $p(\vec{x})$ and the conditional distribution $P(y|\vec{x})$ as varying in time. In particular, a covariate shift, meaning a change in $p(\vec{x})$ over time while $P(y|\vec{x})$ stays constant, is called *virtual concept drift*. A change in $P(y|\vec{x})$ over time is called *real concept drift*. Prior research in concept drift has focused on either adapting a model over time to smooth and slow concept drifts or detecting a point of sudden concept drift, such that the old model can be discarded and a new model can be learned [2]. Recently, explicit long and short term memory models demonstrated an excellent ability to cope with different types of concept drift [8].

Our example of electrode shifts in bionic hand prostheses is best described by a sudden, real concept drift, in which case concept drift theory would recommend to discard the existing classifier and re-train a new one [2]. However, re-learning a viable classifier model may require considerable amounts of new training data to be recorded, which is inconvenient or even infeasible in user's everyday lives. Instead, we would like to re-use an existing classifier model and *adapt* it to the disturbed situation. This approach is motivated by prior research on myoelectric data which indicates that disturbances to electrode shifts are typically simple in structure, that is, they tend to be signal amplitude changes and shifts in the frequency spectrum [6]. Therefore, learning to transfer between the disturbed and the undisturbed setting may be considerably simpler compared to learning a new model [9].

Learning such transfers between domains has been studied in the fields of *domain adaptation* and *transfer learning*. Domain adaptation refers to re-using an existing model in another domain where little to none new training samples are available [3]. Similarly, transfer learning refers to the transfer of knowledge from a source domain, where a viable model is available, to a target domain, where the prior and/or conditional distribution is different [10]. In particular, rather than adjusting the probability distribution in a given data space, transfer learning focuses on adapting the data *representation*. Conceptually, this fits well to our setting as the data representation in terms of EMG readings changes, while the underlying data source, i.e. the neural code of the desired motion, remains the same.

Our key contribution is an efficient algorithm for transfer learning on labeled Gaussian mixture models relying on expectation maximization [11]. In particular, we learn a linear transformation which maps the target space training data to the source space such that the likelihood of the target space data according to the source space model is maximized. This approach generalizes to discriminative models, in particular learning vector quantization models, such as generalized matrix learning vector quantization (GMLVQ), or its localized version, LGMLVQ [12]. We evaluate our approach on artificial as well as real myoelectric data and show that our transfer learning approach can learn a transfer mapping which improves classification accuracy significantly and outperforms all tested baselines, if few samples from the target space are available and/or these samples do not cover all classes.

We begin by discussing related work, continue by introducing our own approach and conclude by evaluating our approach in comparisons to baselines from the literature.

2. Related work

We begin our comparison to related work by introducing some key concepts of transfer learning more formally. In our setting, we assume that a classification model $f : \mathcal{X} \to \{1, \ldots, L\}$ has been trained in some source space $\mathcal{X} = \mathbb{R}^m$ for some $m \in \mathbb{N}$ and we want to apply this model f in some target space $\hat{\mathcal{X}} = \mathbb{R}^n$ for some $n \in \mathbb{N}$. Note that we assume that the classification task itself is the same for both spaces. This makes our setup an instance of *domain adaptation* [3] or *transductive transfer learning* [10]. In the example of an electrode shift on EMG data, we have $m = n$, but a simple application of our source space classifier f is hindered by a fact that the activation pattern is rotated in the feature space and thus the joint distribution $p_{\hat{\mathcal{X}}}(\hat{x}, \hat{y})$ for data $\hat{x} \in \hat{\mathcal{X}}$ and labels $\hat{y} \in \{1, \ldots, L\}$ in the target space differs from the joint distribution in the source space $p_{\mathcal{X}}(\vec{x}, y)$ (see Fig. 1).

One family of approaches to address domain adaptation problems are importance sampling approaches, such as kernel mean matching [13], which apply a weight to each data point in the source space and re-learn the model f with these weighted data points in order to generalize better to the target space [10]. The weights approximate the fraction $\frac{p_{\hat{\mathcal{X}}}(\vec{x})}{p_{\mathcal{X}}(\vec{x})}$, that is, the proportion of the probability of a point in the source space and in the target space. It can be shown that these weights minimize the empirical risk in the target space, if the conditional distributions in both spaces are equal, that is, $p_{\hat{\mathcal{X}}}(y|\vec{x}) = p_{\mathcal{X}}(y|\vec{x})$ [10]. However, this rather demanding assumption does not hold in our case because

Please cite this article as: B. Paaßen et al., Expectation maximization transfer learning and its application for bionic hand prostheses, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2017.11.072