ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Similarity-Based Zero-Shot Domain Adaptation for Wearables

Markus Vieth[1], Nils Grimmelsmann[2], Axel Schneider[2], and Barbara Hammer[1] *

1- Bielefeld University - Machine Learning Group
Inspiration 1, 33619 Bielefeld - Germany

2- Hochschule Bielefeld - University of Applied Sciences and Arts
Interaktion 1, 33619 Bielefeld - Germany

Abstract. Biosensors measure signals from the human body, and usually process them with a small ML model on simple hardware. When a new person starts using such a device, a domain adaptation problem arises. We consider the case where no labels are known for the new person, but data (including labels) from several other people are available (unsupervised, multi-source). As an application scenario, we look at a shoe insole with 3-8 pressure sensors that estimates how much weight/force is put on the foot (regression problem). We propose a distance measure between a source and target domain, and a combination of all source models. Experiments on real world data from 13 persons show that our method outperforms all other tested methods by a good margin.

1 Introduction

Wearable sensors that aid in diagnosis and post-surgery care are becoming more common [3]. In this paper, we focus on a shoe insole equipped with several pressure sensors that can compute the weight put on the foot/leg (e.g. while walking) and warn if it is overstrained, an important application after femur fractur [1]. This is one example of soft-sensors, i.e., hardware components together with subsequent intelligent and adaptive post-processing [4].

A relevant problem in this context is how to adapt the prediction when a new user starts using the device. This should be possible adjusting the software component only but not the hardware, i.e., it can be modeled as domain adaptation (DA) problem [2]. The following characteristics hold: sensor inputs and target values are available for multiple people from historic data. For the target person, only sensor inputs but no real-valued target values are available, i.e., zero-shot learning strategies for regression become necessary [9]. All methods need to be implementable on the edge, working with low computational power.

One challenge which occurs in this context is the high degree of individuality of persons: this concerns different weights and different geometry of feet between persons, and different walking style and duration of the stance phase even for single persons. This makes robust domain adaptation methods necessary.

*The project was funded by the Ministry of Culture and Science of the Federal State North Rhine-Westphalia in the frame of the project RoSe in the AI-graduate-school `dataninja.nrw`, and the project SAIL, grant ID NW21-059A, `sail.nrw`.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

In this work, we present a domain adaptation method that enables the transfer of models to other persons without any given label information, and that compares favorably on a realistic benchmark data set.

2 Related Work

Many domain adaptation methods have been proposed, differing in whether they need *labeled* target data, and in their assumed difference between source and target domains. For classification tasks, domain adaptation often relies on clustering of the data distribution or class-wise transfers. Such principles are not applicable for regression tasks, and domain adaptation for regression problems is less common than for classification problems. Here we survey relevant methods:

Subspace Alignment (SA, [6]) first computes a Principal Component Analysis on the source and the target domain; these are matched via a linear transformation, and a classifier or regressor, trained on the labeled source data, is transformed to the target subspace.

Correlation Alignment (CORAL, [10]) determines the covariance matrices of source and target data, and then computes a linear transform from the source covariance to the target covariance. The classifier/regressor is again trained on the source data, transformed to the target domain.

Transfer Component Analysis (TCA, [8]) computes a kernel matrix of all possible sample pairs (source-source, target-target, and source-target), and then computes a linear mapping to a lower dimensional space via an Eigen decomposition. For training a model, the labeled source data is transformed to this space. Evaluating the model involves computing the kernel on each pair of testing data and source plus target data.

In Domain-Adversarial Neural Networks (DANN, [7]) the idea is to have a feature extractor, trained so that it is impossible to determine whether its output came from the source or the target domain. A domain classifier is introduced, trying to distinguish between the domains in the feature extractor output, and both are trained in an adversarial fashion. Additionally, a label predictor takes the feature extractor output and is trained on the labeled source data. Extensions to multi-source problems and regression problems exist, such as multisource domain adversarial networks (MDAN, [13]). DANN and MDAN have the disadvantage that training the feature extractor, domain classifier, and label predictor is computationally rather expensive.

Some multi-source DA methods combine models from the source domains. BayesMSDA [11] follows the Bayesian principle of determining the posterior via the prior and the likelihood: the prior is computed for each source domain, based on whether the classifier from that domain gives a similar output for two similar target samples (sample similarity is defined by L^1 distance). The likelihood for a target sample is the mean distance of the sample to its K nearest neighbors from the source domain S_m. Finally, the source classifiers are weighted proportionally to the product of prior and likelihood. Since the likelihood is defined per sample, this has to be calculated for every sample (including the nearest neighbor search

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2.
Available from http://www.i6doc.com/en/.

on each source domain), which makes an evaluation rather costly.

In the following, we will compare our model to the DA frameworks which lend itself to edge computing, i.e., SA, CORAL, and TCA; further BayesMSDA.

3 Zero-shot domain adaptation based on data similarities

We model zero-shot learning of the soft-sensor for one person from other persons as a DA problem, where different persons correspond to different domains. The DA method we propose can be divided into two parts: computing the similarity of domains via a distance function, and a weighted average of all source domains where weights are controlled by the similarity.

Generally, data consist of time series of sensor values X_s of n_f sensors and n_s steps/stance phases, taken from a person's walk over a pressure mat. In addition, label information y_s corresponding to the weight put on a foot is available for source domains. First we describe how to estimate the similarity between a source domain s_i and the target domain t. The reasoning for the subsequent similarity measure is the following: the samples where each sensor has the maximum value are easy to determine, and still carry a lot of information. The assumption is that, for two similar domains, the desired prediction at a sensor maximum is similar, and the same model can make that prediction.

We assume as given: source data X_{s_i}, corresponding source labels y_{s_i}, target data X_t, body weight of source and target person bw_{s_i}, bw_t, and a prediction model $predict_{s_i}(x)$, trained on source data and labels. For each sensor and step, we determine which sample has the highest value at this sensor, and store it. Notice that the data can be treated as a stream, without storing all samples, by replacing samples if one with a higher value at the specific sensor was found, and by recognizing that a step has ended if all sensors measure (close to) zero pressure. This leaves us with $n_f \cdot n_s$ samples. Next we apply the source model on all samples, divide by body weight, and compute the mean over all n_s steps, for each sensor j individually

$$\hat{y}_{t,j} = \frac{1}{bw_t n_s} \sum_{k=1}^{n_s} predict_{s_i}(x_{t,j,k})$$

(same for $\hat{y}_{s_i,j}$). Then we compare the predicted values of the source and the target domain, per sensor: if one is more than double of the other, we dismiss the pair, assuming that it is unfit for comparison. We compute the mean squared difference for all remaining pairs leading to the difference of two domains,

$$d(s_i, t) = mean(\{(\hat{y}_{s_i,j} - \hat{y}_{t,j})^2 | 0.5\hat{y}_{s_i,j} < \hat{y}_{t,j} < 2\hat{y}_{s_i,j}, j = 1, \ldots, n_f\}).$$

After computing the domain differences between the target domain and all source domains, we transform these distances into weights for a weighted average. A small distance should translate to a large weight and vice-versa, so we use the exponential function on the negative distance:

$$w_{s_i} = exp(-d(s_i, t)/median(d(s_1, t), d(s_2, t), ...))$$

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

We normalize the weights to sum to 1. The prediction for a target sample x is

$$predict_t(x) = \sum_{s_i} w_{s_i} predict_{s_i}(x).$$

We refer to this domain adaptation based on similarities as OURS.

4 Experiments

For our experiments, we use the dataset from [12]: this consists of data from 13 persons walking over a high resolution pressure sensor array, 10 steps ($n_s = 10$) from each foot. The median number of samples per foot is 453.5. The sensor placement/selection is computed with the best performing method from [12].

We compare the following methods:

SA, CORAL, TCA: for TCA we use a linear kernel; RBF and polynomial kernels lead to a negative R^2 score and were thus not further considered

BayesMSDA: using linear regression for each source domain

OSNA: one source no adaptation – a regression model trained on one random source domain, without adaptation to the target domain, as a lower bound

MSNA: multi source no adaptation – a regression model trained on data from all source domains, without adaptation to the target domain. The fact that the training data comes from multiple domains is not taken into account. This is another, stronger baseline.

TARGET: a regression model is trained on the data from the target domain (via cross-validation). Note that in reality this is not a feasible method since the training needs *labeled* target data, which we assumed is unavailable in practice. This model serves as an upper baseline.

OURS: our proposed method. Here, all reported scores are 5-fold crossvalidation scores, meaning that 4 parts of the target data are used to determine the target model, and the fifth part for evaluation, averaged over five runs.

We use SA, CORAL, and TCA from the ADAPT library [5]. Hyperparameters are optimized and set as follows: for SA, the number of components is equal to the number of features (i.e. the number of sensors). Since for CORAL, hyperparameter optimization sets the regularization parameter very large, which would imply no adaptation at all, we instead choose it as 1 as proposed in the original paper. For TCA we set $\mu = 0.6128$ and the number of components to 7.

We perform two variations of the experimental evaluation: Firstly, we keep the number of source persons fixed at 10 and vary the number of sensors n_f between 3, 5, and 8. This way, we test how well the DA methods deal with more or less complex data. Secondly, we set the number of sensors to 5, and use 3, 6, or 10 source persons to examine how the performance of the multisource

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Table 1: Varying number of sensors/features, R^2 score (higher values are better), mean ± standard deviation, 10 source persons. Note that TARGET is the upper baseline and has the target labels, unlike the other methods. "cost" is relative to TARGET, approximated from the runtime on a CPU, and refers to 8 sensors.

method	3 sensors		5 sensors		8 sensors		cost
OSNA	0.740	± 0.248	0.843	± 0.150	0.864	± 0.211	1.0
SA	0.697	± 0.197	0.807	± 0.141	0.875	± 0.119	2.5
CORAL	0.635	± 0.370	0.654	± 0.335	0.570	± 0.462	1.9
TCA	0.739	± 0.248	0.843	± 0.150	0.855	± 0.215	656
MSNA	0.825	± 0.117	0.913	± 0.069	0.942	± 0.051	1.2
BayesMSDA	0.829	± 0.111	0.914	± 0.070	0.938	± 0.054	2894
OURS	0.866	± 0.081	0.926	± 0.055	0.954	± 0.036	8.5
TARGET	0.901	± 0.049	0.948	± 0.033	0.978	± 0.013	1.0

methods changes. For comparison, the coefficient of determination (R^2) is used, which is higher the better the prediction is, and 1 for a perfect prediction.

We test 39 different sensor placements (each based on 10 randomly chosen persons), then the domain adaptation is done six times, with each of the feet not used for sensor placement as the transfer target. Thus, each reported mean is from 234 runs.

5 Results

The results for varying the number of sensors are shown in Table 1. SA, CORAL, and TCA perform no better than OSNA (often worse), and BayesMSDA performs similar to MSNA. Our method performs much better than MSNA and BayesMSDA.

Table 2 depicts the results for more or fewer source domains/persons. Even with only 3 source persons, all multisource methods have a notable advantage over single source methods. As it can be expected, all multisource methods become better with more source domains. Our proposed method especially benefits from more source domains: for 3 persons, the difference between our method and MSNA is small, but becomes much larger for 6 and 10 persons.

6 Conclusion

In this work, we have presented a new method for unsupervised multi-source domain adaptation for regression problems. It is based on a problem-specific distance function that respects the unique challenges of this problem. In experiments, we have shown that it performs well, achieving a higher score than all competitors and not far off from the upper baseline. In the future, we might explore how this idea could be used on related domain adaptation problems for wearables.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Table 2: Varying number of source persons, R^2 score (higher values are better), mean ± standard deviation, 5 sensors. Note that TARGET is the upper baseline and has the target labels, unlike the other methods.

method	3 persons		6 persons		10 persons	
OSNA	0.849	± 0.130	0.858	± 0.116	0.843	± 0.150
SA	0.821	± 0.136	0.812	± 0.146	0.807	± 0.141
CORAL	0.628	± 0.485	0.648	± 0.442	0.654	± 0.335
TCA	0.848	± 0.131	0.858	± 0.116	0.843	± 0.150
MSNA	0.908	± 0.069	0.911	± 0.070	0.913	± 0.069
BayesMSDA	0.904	± 0.070	0.910	± 0.072	0.914	± 0.070
OURS	0.912	± 0.064	0.921	± 0.059	0.926	± 0.055
TARGET	0.948	± 0.033	0.948	± 0.033	0.948	± 0.033

References

[1] K. Agarwal-Harding, J. Meara, S. Greenberg, L. Hagander, and G. Dyer. Estimating the global incidence of femur fracture. In *APHA Annual Meeting and Exposition*, 11 2013.

[2] A. Ajith and G. Gopakumar. Domain adaptation: A survey. In M. Tistarelli, S. R. Dubey, S. K. Singh, and X. Jiang, editors, *Computer Vision and Machine Intelligence*, pages 591–602, Singapore, 2023. Springer Nature Singapore.

[3] T. Amin, R. J. Mobbs, N. Mostafa, L. W. Sy, and W. J. Choy. Wearable devices for patient monitoring in the early postoperative period: a literature review. July 2021.

[4] F. Curreri, L. Patanè, and M. G. Xibilia. Soft sensor transferability: A survey, 2021.

[5] A. de Mathelin, F. Deheeger, G. Richard, M. Mougeot, and N. Vayatis. ADAPT: Awesome Domain Adaptation Python Toolbox. *arXiv preprint arXiv:2107.03049*, 2021.

[6] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised Visual Domain Adaptation Using Subspace Alignment. In *2013 IEEE International Conference on Computer Vision*, pages 2960–2967, Sydney, Australia, Dec. 2013. IEEE.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky. Domain-Adversarial Training of Neural Networks, May 2016.

[8] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, Feb. 2011.

[9] Y. Song, T. Wang, P. Cai, S. K. Mondal, and J. P. Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. jul 2023.

[10] B. Sun, J. Feng, and K. Saenko. Return of Frustratingly Easy Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), Mar. 2016.

[11] S.-L. Sun and H.-L. Shi. Bayesian multi-source domain adaptation. In *2013 International Conference on Machine Learning and Cybernetics*, Tianjin, China, July 2013. IEEE.

[12] M. Vieth, N. Grimmelsmann, A. Schneider, and B. Hammer. Efficient Sensor Selection for Individualized Prediction Based on Biosignals. In H. Yin, D. Camacho, and P. Tino, editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2022*, pages 326–337, Cham, 2022. Springer International Publishing.

[13] H. Zhao, S. Zhang, G. Wu, J. M. F. Moura, J. P. Costeira, and G. J. Gordon. Adversarial Multiple Source Domain Adaptation. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.