Transfer Learning Approach to Debt Portfolio Appraisal

Tomasz Kajdanowicz, Slawomir Plamowski, Przemyslaw Kazienko, and Wojciech Indyk

Wroclaw University of Technology, Wroclaw, Poland
Faculty of Computer Science and Management
{tomasz.kajdanowicz,slawomir.plamowski, kazienko,wojciech.indyk}@pwr.wroc.pl

Abstract. Machine learning and data mining algorithms usually assume that the training and future data have the same distribution and come from the same feature space. However, in majority of real-world problems, this is not true. In case of Debt portfolio appraisal we have sufficient training data only in another domain of interest, namely in other portfolios. Therefore, only knowledge transfer from these portfolios in inference for new one is possible. In the paper we propose transfer learning and learning based on similarity methods, basing on similarity between training and testing datasets. The proposed approach is examined in real domain debt portfolio valuation.

Keywords: Transfer learning, dataset selection, distance measures, debt valuation, prediction, supervised learning.

1 Introduction

Supervised learning task is mainly focused on providing inference abilities derived from training data. Training data consist of a set of training examples, each composed of a pair of input features X and a desired output value Y. Therefore the main task is to analyse training data and produce an inferred function Φ that maps input to output, $\Phi : X \rightarrow Y$. If the output is discrete, the function Φ is called a classifier. Otherwise, in case of continuous output, it is called a regression. If function Φ maps to interrelated set of more than one values it is structured prediction or structured output learning algorithm. On the whole, the inferred function Φ should be able to predict the correct output value for any valid input object. This requires learning algorithm to be able to generalize basing on the training data.

However, as it may be expected, the availability of training datasets utilized in learning algorithms has a great influence on the generalization abilities. Sometimes in many real world applications, it is very expensive or even impossible to collect the needed training data to build the models. The traditional inference, based on previous learning in the same domain, is insufficient and in such case it is expected to use more sophisticated knowledge transfer or transfer learning between distinct tasks from similar domains.

E. Corchado et al. (Eds.): HAIS 2012, Part II, LNCS 7209, pp. 46–55, 2012.

The straightforward learning situation arises when learning concerns data from particular domain and describes always the same stationary object. The statistical dependencies between examples remain unchanged and training may be performed using the same source of training and testing data. Such data, as long as being of appropriate size, may deliver satisfactory generalisation abilities and no transfer learning is required. But in order to generalise from data describing non-stationary objects, learning algorithms are expected to model concept drift [7] phenomenon identified by changes in data probability distributions. As concept drift may be caused by changes of prior, conditional or posterior probabilities of data, appropriate methods must address the problem, among them these based on appropriate training set selection.

Another situation occurs when generalisation needs to be performed for objects for which training data is not available. In such case, learning is performed using data from the same domain but describing other similar objects. An example of such a situation are across-network classification where learning performed on one network adjust models used in generalisation on another network [8] or debt portfolio value prediction where value of appraisal of particular portfolio is done using other similar portfolios [5].

The paper considers the problem of transfer learning in the prediction task when inference is based on models learnt on data from this same domain but describing other similar objects. The paper presents a comparison between two learning techniques for that task: learning based on similarity and transfer learning.

Obviously, the greater similarity/smaller distance between objects used in learning and those the inference is applied to, the better performance of inference methods. Similarity/distance identification between training and testing objects can be reduced to similarity/distance measurement between datasets describing their input features, namely similarity/distance between X_{train} and X_{test}. Aforementioned similarity and distance can be invoked interchangeably as similarity can be measured by distance, i.e. two objects are similar if the distance is close to zero. In general, distance is defined as a quantitative degree of how far apart two objects are [2]. The choice of distance measure depends on the representation of objects and type of measurement. Training sets in supervised learning are usually represented by matrices in which columns denote attributes and rows - object instances. A single cell of such matrix contains a value of particular attribute for a given instance. Hence, the problem of learning based on similarity denotes a learning on selected training datasets based on measuring the distance between them and is actually a matrix distance based selection.

On the other hand, transfer learning provides additional ability to apply knowledge derived from external to current datasets for generalisation. The main concern denotes then discovering which knowledge can be transferred and how the knowledge from distinct models should be transferred across domains.

The rest of the paper is organised as follows. In section 2 various approaches of transfer learning and learning based on distance measures are enumerated. In order to provide a better perspective on the considered application problem, section 3 presents a real-world transfer learning problem in debt portfolio value prediction. In section 4 two approaches to transfer learning and learning based on similarity are described. Evaluation of the impact on prediction accuracy using proposed methods is presented in section 5. Finally, section 6 summarises this work.

2 Related Work

In general, the bunch of methods, called here learning based on similarity, assumes that learning a generalisation model is done on training datasets of the same domain. These datasets are selected among all available domains. Training set selection from a set of available historical datasets based on the distance between particular testing set and training set may be considered using two equivalent approaches: as selection based on distance between matrices of non-equal size or, better, as calculating measure of goodness of fit between probability density functions. While calculation of probability density for discrete random variables is performed with respect to the counting measure over the sample space, the density of continuous random variables is given by the integral of this variable's density. This may imply problems as the exact density is not known and the empirical one can be obtained only. Literature proposes either the estimation of probability density function [11] or, simply, consideration of discrete and finite histogram of random variable [2,14]. The histogram can be considered then as a vector, i.e. coordinates in some space, and numerous distances proposed in the literature can be applied to compare two densities.

There exist a substantial number of distance measures derived from various fields such as computer science, information theory, mathematics, physics, or statistics, etc. Some of them that may be used in distance calculation are standard Euclidean distance and KullbackLeibler distance. For a v and w, a vector version of probability density functions of V and W matrices, with length of both vectors equal to d, Euclidean and KullbackLeibler distances are define as in equation 1 and 2, respectively.

$$dist(v,w) = \sqrt{\sum_{i=1}^{d} |v_i - w_i|^2} \tag{1}$$

In general, Euclidean distance measures shortest distance between two points as a length of line and belongs to L_p Minkowski family of distance measures. Applying Shannons concept of probabilistic uncertainty (entropy) Kullback-Leibler distance introduces the relative entropy, called information deviation [4], see equation 2.

$$dist(v,w) = \sum_{i=1}^{d} v_i \ln \frac{v_i}{w_i} \tag{2}$$

Obviously, distance measures presented above are only example ones and a proper choice of representative distance measure depends on the type of measured data and the measurement itself. For further list of distance measures please refer to [2,15].

The approach of calculating distances between vector version of probability density functions tends to be reasonable but requires estimation of probability density function and sometimes it might be troublesome.

The distance of two datasets might be computed by application of other concept - distance matrices. However, this is limited to situation when both datasets have the same size (number of rows) and, what is more, the mapping bijection that states the clear relation of corresponding data examples is known. As the size of compared distinct datasets may differ and the mapping between data examples is not known, it is not always possible to compute distance matrices.

Nevertheless, the distance between datasets may be calculated using matrix norms [9]. The matrix norms are defined in terms of well known vector norms and therefore, it can be said, they are induced by vector norms [16]. As some basic norms like (for a given matrix A) matrix 1-norm - returns maximum of A column sums, matrix ∞-norm - returns maximum of A row sums or matrix 2-norm - returns square root of largest eigenvalue of $A \times A$, more sophisticated once needs to be applied to characterize the matrix [9]. One of them can be Frobenius norm. This norm is the sum of the squares of the Euclidean norms of the matrix columns [9]. Thus it is able to model variability of the data. Investigating the literature we can see that norms are not the perfect solution to model distances between matrices.

On the other hand, transfer learning provides additional ability to learning system making it possible to recognize and utilize knowledge learned in previous tasks (datasets) to new tasks [10]. By that it is meant that transfer learning aims to extract the knowledge from several source tasks and apply the knowledge to a target task. In contrast to previously mentioned learning based on similarity, rather than learning models individually on selected datasets (tasks), transfer learning applies generalised knowledge of all known tasks obtained in single run learning. Figure 1 shows the difference between the learning in traditional and transfer learning approaches. As we can see, the first technique try to learn each task from scratch, while transfer learning utilizes knowledge from some previous tasks to perform an inference.

3 Debt Portfolio Value Prediction

Determining the value of debt portfolios and choosing those with the greatest revenue potential is of a great importance for debt traders. Economically crucial decisions are based on the amount of possible repayment of liabilities. As traders (both buyers and sellers) apply distinct collection processes, amount of receivables obtained may be different. This constitutes the area for trading and to establish a transaction price. Therefore debt portfolio assessment is a complex task. However, as far as machine learning is concerned, this problem may be understood as a prediction task that assesses the possible repayment value from all

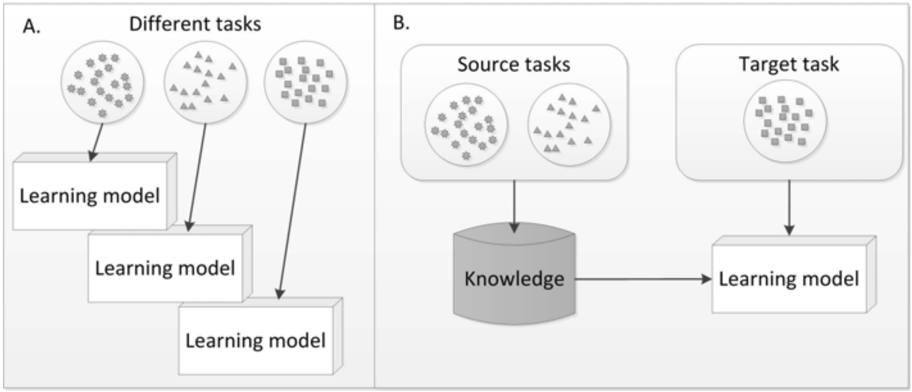

Fig. 1. Different Learning approaches: A – traditional machine learning, B – transfer learning

debt cases belonging to particular portfolio. The repayment is calculated based on historical data of debts.

The most common routine of debt portfolio trade starts when a seller, usually a bank, telecommunication company, etc. offers a set of debts, called debt package or portfolio, expecting a purchase proposal from buyers. Purchasers, usually a specialized debt recovery entities, offer price and the most suitable offer is chosen. The price proposed by a particular buyer may be obtained in variety of ways, among which the utilization of historical data of debt recovery in order to build a prediction model seems the most reasonable one. Such model provides an estimation of possible return from the package.

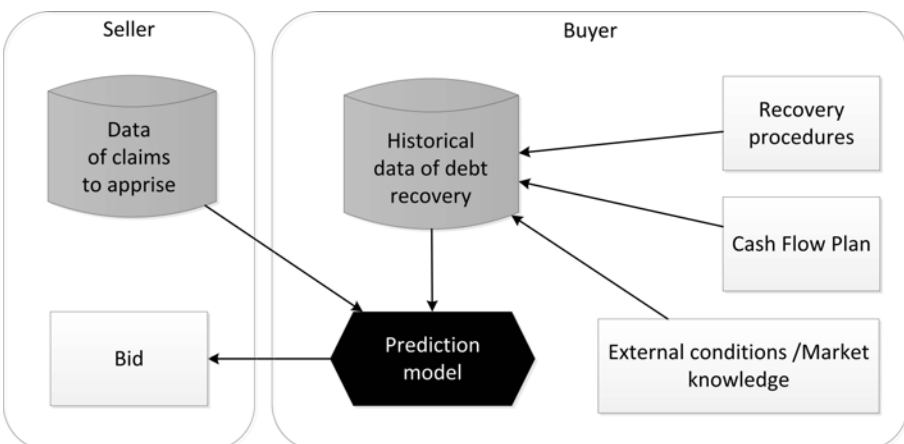

Fig. 2. The process of debt portfolio purchase with utilization of prediction model

In considered situation the valuation of debt portfolio is based on the data of historical claims with their repayment profiles over time. A debt collection company usually assumes that gathered repayment data reflects all important dependencies influencing repayment results like recovery procedures, cash flow plans and other external conditions. Such assumption simplifies the problem as changes in the probabilities caused by evolving business environment are ignored. The model trained on historical data is applied to predict the repayment amount of the offered portfolio. Based on the obtained results, bids are offered to the seller. The process of debt portfolio valuation for bid proposal is presented in Figure 2.

Summarizing, the most significant and sensitive part of debt trade is repayment value prediction process. The accuracy of prediction for offered portfolio relies mainly on model generalization capabilities and quality of training data. As it is very difficult to provide prediction using whole, large historical data, some training dataset selection mechanism needs to be employed. In the further part of the paper we present the method for training set selection for model learning, that is applied to considered business scenario.

4 Learning Based on Similarity and Transfer Learning Techniques for Debt Portfolio Appraisal

Assuming that training set can be treated as a matrix, the problem of training set selection is equivalent to the matrix selection using some notion of distance. Assuming additionally that the environment remains stationary, the generalisation can be done on the basis of historical datasets of the same domain, which describe similar objects with the same attributes. In the aforementioned debt prediction problem, historical dataset consist of debt portfolios, that have already been repaid. They are used to predict the repayment value of unknown, new portfolio. The learning could be done using all historical datasets, but from the practical point of view it would not always be possible (e.g. massive training data) and of high quality (poor inference from complex and non-discriminative data). Therefore some hybrid learning methods need to be applied.

Hereby we propose a method based on transfer learning concept, which one can describe as using knowledge from previous prediction tasks to acquire new knowledge in current task. One can make an assumption that aforetime acquired knowledge can be not only helpful, but also essential for future prediction. Such consideration can provide additional latent information that are transferred during training process.

The proposed method is based on the assumption that there exists a set T of k train sets A_i, $i \in \{1, \ldots, k\}$, $k \in \mathbb{N}$ and a single test set B. The actual task is to create a ranking of distances between each set A_i from T and test set B. Having created such ranking, train sets A_i are sorted in ascending order of distances. In the next step, the method of learning based on similarity utilizes closest A_i

sets for training procedure, whereas transfer learning generalizes all datasets and from now one is able to transfer the knowledge to new tasks, see differences in algorithms 1 and 2. It means that inference does not require training the models. Each dataset is generalised by one learning model and in order to balance consume the knowledge from multiple tasks, weights vector is calculated as sum of distances between considered train sets A_i and test set B, namely $dist(A_i, B)$, divided by the sum of distances, see equation 3. Trained predictors are then used to obtain results from testing set B (algorithm 2).

$$weight\_vector_i = \frac{dist(A_i, B)}{\sum_{k=1}^{M} dist(A_k, B)}, \tag{3}$$

In proposed method, separate prediction algorithm is used for each train set A_i. Afterwards we use these predictors to infer targets on test set B. Results of inference from different tasks are weighted by the aforementioned weights vector.

Algorithm 1. The pseudo code of learning and inference phase of the method for learning based on similarity

Require: set T of k training sets A_i, $i \in \{1, \ldots, k\}$, testing set B

```
for all training sets A_i ∈ T do
    calculate dist(B, A_i)
end for
build a distance ranking
select training dataset(s) using ranking
build model on selected dataset(s)
return inferred targets for B
```

Algorithm 2. The pseudo code of learning phase of the transfer learning approach

Require: set T of k training sets A_i, $i \in \{1, \ldots, k\}$

```
for all training sets A_i ∈ T do
    learn the model
end for
return set of models
```

5 Experiments and Results

The main objective of performed experiments was to test and evaluate the proposed transfer learning technique in debt appraisal task. Among others some standard performance measures were observed: Relative Error(RE), Mean Square Error (MSE), Coefficient of Correlation (R), Variance Accounted For (VAF), Maximum Absolute Error (MAE), Coefficient of Efficiency (COE).

Algorithm 3. The pseudo code of inference phase of the transfer learning approach

Require: set T of k training sets A_i, $i \in \{1, \ldots, k\}$, testing set B, set of learnt models Φ

```
1: for all training sets A_i ∈ T do
2:     calculate dist(B, A_i) and weight_vector_i
3: end for
4: infer targets for B using Φ and weight_vector_i
5: return inferred targets for B
```

Experiments were carried out on fifteen distinct, real datasets from the same application domain of debt portfolio pattern recognition [5]. Datasets represent the problem of aggregated prediction of sequential repayment values over time for a set of claims.

The procedure of experiment accomplishes a prediction of possible repayment values for a B debt portfolio. Depending on learning approach, from among all or selected known portfolios learning sets are constructed. Using selected packages, the regression algorithms are trained and eventually basic tests for portfolio B are performed.

Based on described procedure, three distinct experimental scenarios are created. They vary in the number of selected portfolios for training and in utilized inference method, namely learning based on training set similarity and transfer learning. Therefore the best known methods from authors' previous findings [6] are compared with the transfer learning approach in the experiments. The first scenario uses the closest package for learning, the second – three closest packages and the third all packages but with distinct inference procedure. From this point, these scenarios are denoted as: *C, C3, TL* respectively. For examined scenarios Friedman test is performed. Results are presented in Table 1.

Table 1. Average rank positions determined in Friedman test

Measure/Rank	1^{st}	2^{nd}	3^{rd}
RE	TL (1.67)	C3 (2.13)	C (2.20)
MSE	C3 (1.47)	C (1.87)	TL (2.67)
R	C (1.53)	TL (2.00)	C3 (2.47)
VAF	TL (1.47)	C (1.93)	C3 (2.6)
MAR	TL (1.67)	C3 (1.93)	C (2.4)
COE	TL (1.33)	C (2.13)	C3 (2.53)

For each prediction algorithm statistical ranking is created to indicate optimal approach. We incorporated Friedman statistical test as intuitive and convenient procedure for different used approaches comparison. Mean rank position for each combination of method and scenario is shown in parentheses. The lower rank value, the lower observed performance measure yielded by prediction process.

As shown in Friedman test, usage of transfer learning approach results in better performance than using single and multiple closest datasets for training.

The results in Table 1 can be read as follows: for fifteen debt evaluation tasks selecting transfer learning approach, denoted by TL, results in the smallest mean squared Relative Error (RE), Variance Accounted For (VAF), Maximum Absolute Error (MAE) and Coefficient of Efficiency (COE). Friedman's test places this learning approach in the first place of ranking.

As it can be observed in Table 1 the TL approach performs worse in Mean Square Error (MSE) and Coefficient of Correlation (R) measures in comparison with other methods. However, according to the nature of debt portfolio evaluations the objective is to minimize the Relative Error (RE). Therefore some approaches may be better in MSE minimization while it is not a main target.

6 Conclusions

The problem of transfer learning was considered in the paper. We introduced a learning based on similarity method, that selects training sets to be used in training. Sets are chosen based on distance between two datasets. Moreover, we proposed transfer learning approach to this same task. Transfer learning method does not require model learning each time the inference needs to be employed.

The proposed methods were examined on real datasets in the debt portfolio valuation domain. The results indicated that proposed transfer learning method can be used to infer effectively in the debt portfolio appraisal domain.

Further experimentation will consider a comparison of the presented method with other approaches. Moreover, further studies will focus on discovery and description of properties of proposed method.

Acknowledgement. This work was partially supported by The Polish Ministry of Science and Higher Education the research project 2011-2012, 2011-2014 and Fellowship co-financed by The European Union within The European Social Fund.

References

1. Cano, J.R., Herrera, F., Lozano, M.: Using evolutionary algorithms as instance selection for data reduction in KDD: an experimental study. IEEE Transactions on Evolutionary Computation 7(6), 561–575 (2003)
2. Cha, S.H.: Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods in Applied Sciences 1(4), 300–307 (2007)
3. Coifman, R.R., Wickerhauser, M.V.: Entropy-based algorithms for best basis selection. IEEE Transactions on Information Theory 38, 713–718 (1992)
4. Deza, E., Deza, M.M.: Dictionary of Distances. Elsevier (2006)
5. Kajdanowicz, T., Kazienko, P.: Prediction of Sequential Values for Debt Recovery. In: Bayro-Corrochano, E., Eklundh, J.-O. (eds.) CIARP 2009. LNCS, vol. 5856, pp. 337–344. Springer, Heidelberg (2009)
6. Kajdanowicz, T., Plamowski, S., Kaznieko, P.: Training Set Selection Using Entropy Based Distance. In: The IEEE Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2011, pp. 340–344. IEEE Computer Society (2011)

7. Kurlej, B., Wozniak, M.: Active learning approach to concept drift problem. Logic Journal of the IGPL (2011), doi:doi:10.1093/jigpal/jzr011
8. Lu, Q., Getoor, L.: Link-based classification using labeled and unlabeled data. In: ICML 2003 Workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining (2003)
9. Meyer, C.D.: Matrix analysis and applied linear algebra. Society for Industrial and Applied Mathematics (2000)
10. Pan, S.J., Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22(10), 1345–1359 (2010)
11. Rencher, A.: Methods of multivariate analysis. John Wiley & Sons (2002)
12. Son, S.-H., Kim, J.-Y.: Data Reduction for Instance-Based Learning using Entropy-Based Partitioning. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3982, pp. 590–599. Springer, Heidelberg (2006)
13. Theodoris, S., Koutroumbas, K.: Pattern Recognition. Elsevier (2009)
14. Toussaint, G.T.: Bibliography on estimation of misclassification. IEEE Transactions on Information Theory 20(4), 472–479 (1974)
15. Ullah, A.: Entropy, divergence and distance measures with econometric applications, Department of Economics. University of California - Riverside (1993)
16. Zhou, K., Doyle, K., Glover, K.: Robust and Optimal Control. Prentice-Hall (1996)