2014 22nd International Conference on Pattern Recognition

Domain Adaptation Under Data Misalignment: An Application to Cepheid Variable Star Classification

Ricardo Vilalta
Department of Computer Science
University of Houston
Houston TX, 77204-3010
Email: vilalta@cs.uh.edu

Kinjal Dhar Gupta
Department of Computer Science
University of Houston
Houston TX, 77204-3010
Email: kinjal13@cs.uh.edu

Lucas Macri
Department of Physics and Astronomy
Texas A&M University
College Station, TX 77843
Email: lmacri@tamu.edu

***Abstract*—We address a particular scenario within the area of domain adaptation, where a predictive model obtained from a source domain can be applied directly to a target domain. Both source and target domains share the same input or feature space, but we do not impose any restrictions on the marginal and class posterior distributions (both distributions can differ). Our main assumption is that the difference between the source and target domains can be traced to a systematic change caused by some modification in the sensing device, or environment surrounding the phenomenon under analysis, as for example when the training and testing samples correspond to stars (i.e., to light curves) that belong to different galaxies. We demonstrate how such systematic change can be reversed by shifting the target data towards the source data until both distributions are aligned. Our approach uses maximum likelihood to compute the right amount of displacement along each variable under analysis. We test our methodology on the classification of Cepheid variable stars according to their pulsation mode: fundamental and first-overtone. Experimental results with three galaxy datasets (Large Magellanic Cloud, Small Magellanic Cloud, and M33), show the effectiveness of our approach.**

I. Introduction

Domain adaptation seeks to adjust the model obtained in a source domain to account for differences exhibited in a new target domain. Unlike traditional studies in classification where both training and testing sets are assumed as realizations of the same joint input-output distribution, domain adaptation either weakens or completely disregards such assumption [1], [2], [3]. In addition, domain adaptation commonly assumes an abundance of labeled examples in the source domain, but little or no class labels in the target domain. An example of these concepts lies in light curve classification from star samples obtained from different galaxies. A classification task set to differentiate different types of stars in a nearby source galaxy –where class labels are available– will experience a change in distribution as we move to a target galaxy lying farther away –where class labels are unavailable; a major reason for such change is that at greater distances, less luminous stars fall below the detection threshold and more luminous stars are preferentially detected. The corresponding dataset shift [4], [5], [6], precludes the direct utilization of one single model across galaxies; it calls for a form of model adaptation to compensate for the change in the data distribution.

Many other domains exist where the difference between the target and source domains are primarily caused by measurements obtained under different circumstances (e.g., observing galaxies that lie at different distances), by changing the orientation or position of the same sensing device, or by utilizing a similar but more powerful device. Our main assumption is that the change between domains is systematic and affects all points similarly. In particle physics, for example, a model built to identify a certain particle is rendered unapplicable when we collect samples obtained using more powerful particle accelerators; this is because the range of parameter values shifts as we reach out to higher energies. The shift, however, is systematic, in that all particle collisions are exposed to the same new energies, and hence we expect to observe a similar displacement across all points in the new target distribution. Other practical examples can be found in medical imaging, where technological advancements keep producing more sophisticated imaging equipment; in radar or sonar signal processing, where instrumentation has improved in providing better data quality, etc.

In this paper, we describe a methodology for domain adaptation that assumes two possible sources of discrepancy between source and target distributions: a data misalignment caused by a systematic change during data collection (e.g., changing the physical conditions surrounding an experiment); and a sample bias where the change in physical conditions gives more weight to some examples or events (i.e., makes some examples or events easier to detect), and thus increases their probability density. We contend that many real-world scenarios exhibit this same type of distribution change. Following these assumptions, our proposed approach to domain adaptation is to shift the target data towards the source data until both distributions look as similar as possible. This implies computing the right amount of displacement along each variable under analysis. We then invoke the original model trained in the source domain, and apply it in a straightforward manner on the transformed (i.e., shifted) target data.

We test our methodology on the classification of Cepheid variable stars [7]; we focus on two subtypes of Cepheids according to their pulsation mode: fundamental and first-overtone. The importance of attaining a robust discrimination between fundamental and first-overtone Cepheids lies in the need for increased precision and accuracy when calculating distances to nearby galaxies. In addition, a robust discrimination between fundamental-mode and first-overtone pulsators can be used to compare the predictions of stellar evolution and pulsation models. These models compute changes in the internal structure and chemical compositions of stars from their initial conditions (main sequence) until the end of the Cepheid phase and beyond. Frequent misclassifications would result in evident flaws in the modeling phase.

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.629

We characterize each Cepheid star using features extracted from its corresponding light curve, and then invoke our methodology to re-use predictive models across galaxy datasets. Different from previous work [8], [9], our study looks into geometric properties along the light curve, that serve as better discriminators between the two types of pulsation mode (fundamental and first-overtone). Empirical results show an increase in predictive accuracy with our proposed methodology.

Some of the challenges addressed in this paper have been attacked before by others [10], [8]; different from previous studies, our work incorporates new light-curve descriptors by attending to geometric properties of an interpolating function across light-curve points. In addition, we extend previous work [10], by generalizing the form of the maximum likelihood estimator to any number of features, and by analyzing scenarios where domain adaptation is expected to be beneficial, through the use of information-theoretic divergence metrics.

The paper organization follows. Section II begins by providing basic concepts in classification and dataset shift. Section III explains our proposed methodology. Section IV described related work. Section V explains our light-curve geometric features. Section VI shows our experiments using real astronomical data. Lastly, Section VII gives a summary and conclusions.

II. Preliminary Concepts in Statistical Learning

Consider an n-component vector-valued random variable $(X_1, X_2, \cdots, X_n)$, where each X_i represents an attribute or feature; the space of all possible feature vectors is called the input space $\mathcal{X}$. Consider also a set $\{w_1, w_2, \cdots, w_k\}$ corresponding to the possible classes; the space of all possible classes is called the output space $\mathcal{W}$. A classifier typically receives as input a set of training examples from a source domain, $T_{\rm tr} = \{(\mathbf{x_i}, w_i)\}_{i=1}^{p}$, where $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ is a vector in the input space, and w is a value in the (discrete) output space. We assume the training or source sample $T_{\rm tr}$ consists of independently and identically distributed (i.i.d.) examples obtained according to a fixed but unknown joint probability distribution, $P_{\rm tr}(\mathbf{x}, w)$, in the input-output space $\mathcal{X} \times \mathcal{W}$. The outcome of the classifier is a hypothesis or function $f(\mathbf{x}|\theta)$ (parameterized by θ) mapping the input space to the output space, $f : \mathcal{X} \rightarrow \mathcal{W}$.

We commonly choose the hypothesis that minimizes the expected value of a loss function $L(w, f(\mathbf{x}|\theta))$, also known as the risk:

$$R(\theta, P(\mathbf{x}, w)) = E_{\sim P}[L(w, f(\mathbf{x}|\theta))] \quad (1)$$

where we typically adopt the zero-one loss function $L(w, f(\mathbf{x}|\theta)) = 1_{\{\mathbf{x}|w \neq f(\mathbf{x}|\theta)\}}(\mathbf{x})$; $1(\cdot)$ is an indicator function.

We will also assume a testing sample from a target domain, $T_{\rm te} = \{\mathbf{x_i}\}_{i=1}^{q}$, that consists of i.i.d. examples obtained from the marginal distribution $P_{\rm te}(\mathbf{x})$ according to a different joint distribution, $P_{\rm te}(\mathbf{x}, w)$, over $\mathcal{X} \times \mathcal{W}$. We will consider the case where originally both source and target domains differ in the marginal distributions $P_{\rm tr}(\mathbf{x}) \neq P_{\rm te}(\mathbf{x})$, and class posteriors (i.e., conditional distributions) $P_{\rm tr}(w|\mathbf{x}) \neq P_{\rm te}(w|\mathbf{x})$. After we have applied our methodology, we will assume that the only difference between the training and test distributions is in the marginal distributions.

III. The Data Alignment Problem

The problem we address is characterized by an original source domain where class labels abound, and by a target domain with no class labels. We show how the target domain can still be learned by assuming the difference in the joint input-output distribution is mainly due to a systematic shift of sample points. We illustrate an example in the context of Cepheid variable star classification, where we wish to discriminate stars according to their pulsation mode(s); specifically, we focus on the two most abundant classes, which pulsate in the fundamental and first-overtone modes. Such classification can in fact be attained for nearby galaxies with high accuracy (e.g., Large Magellanic Cloud) under the assumption of class-label availability. The high cost of manually labeling variable stars, however, suggests a different mode of operation where a predictive model obtained on a data set from a source galaxy $T_{\rm tr}$, is later used on a test set from a target galaxy $T_{\rm te}$. Such scenario is not straightforwardly attained, as shown in figure 1 (left), where the distribution of Cepheids in the Large Magellanic Cloud LMC galaxy (source domain, top sample), deviates significantly from that of M33 galaxy (target domain, bottom sample). In this example, we employ two features only: apparent magnitude in the y-axis, and log period in the x-axis, but our solution is general and allows for a multi-variate representation. Both the offset along apparent magnitude[1], and the significant degree of sample bias, are mostly due to the fact that M33 is $\sim 16\times$ farther than the LMC. At these greater distances, the shorter-period (i.e., less luminous) Cepheids fall below the detection threshold and longer-period (i.e., more luminous) stars are preferentially detected. This is what we refer to as a systematic shift, where all objects or events (i.e., sample points) are exposed to the same change, and thus we expect a similar displacement throughout the entire target distribution.

Our general solution is to shift the data on the target domain, $T_{\rm te}$, to the point where the model previously obtained over the source domain, $T_{\rm tr}$, can be re-used. We call this the *data-alignment problem* [10]. This differs from previous methodologies where the goal is to generate a model afresh by re-weighting examples on the training set [4]; instead we leave the model intact, and transform the testing set until the model can be re-used. An example of a successful alignment is shown in Fig. 2 (right), where shifting the sample of M33 along the y-axis (apparent mean magnitude) by the right amount, enables us to do classification with the same model obtained on the source domain.

A. Problem Formulation.

We generalize our problem as follows. Assume a training set $T_{\rm tr}$ and a predictive model $f(\mathbf{x}|\theta)$ obtained by training on $T_{\rm tr}$. Assume also an unlabeled testing set $T_{\rm te} = \{\mathbf{x}\}$, where each feature vector can be split into two parts: $\mathbf{x} =$

[1]The flux of each star is represented by its apparent magnitude m, defined as $m = -2.5 \times \log_{10} \frac{L}{d^2}$, where d is the distance from Earth to the star measured in parsecs, and L is the star luminosity. Hence, smaller numbers correspond to brighter magnitudes (higher fluxes).

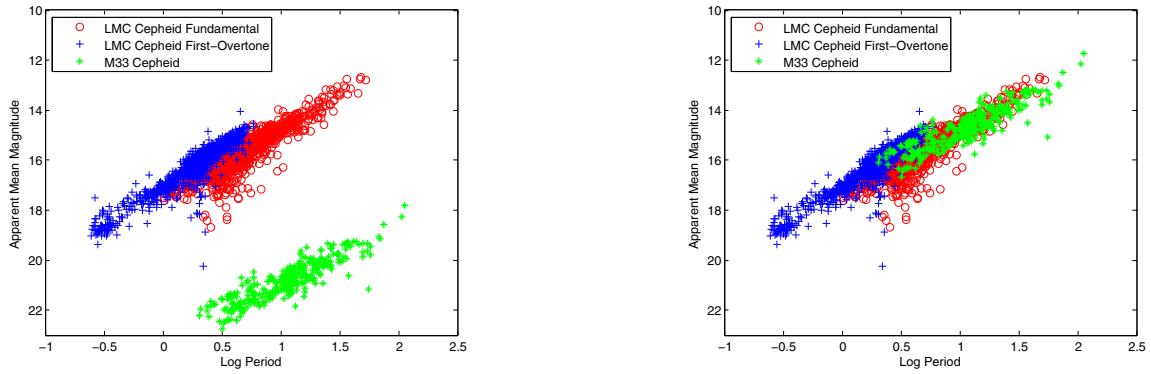

Fig. 1. Left. The distribution of Cepheids along the Large Magellanic Cloud LMC (top sample), deviates significantly from M33 (bottom sample). Right. M33 is aligned with LMC by shifting along mean magnitude.

$(X_A, X_B) = (x_1, x_2, ..., x_m, x_{m+1}, x_{m+2}, ..., x_n)$. Without loss of generalization, we assume each feature in the set $X_B = \{x_{m+1}, x_{m+2}, ..., x_n\}$ has suffered a shift. We wish to generate a new dataset $T'_{\text{te}} = \{x'\}$, where $\mathbf{x}' = (X_A, X'_B) = (x_1, x_2, ..., x_m, x_{m+1} + \delta_{m+1}, x_{m+2} + \delta_{m+2}, ..., x_n + \delta_n)$, to achieve the goal of aligning T_{te} along T_{tr}. Our real focus is in the set $\Delta = \{\delta_{m+1}, \delta_{m+2}, ..., \delta_n\}$; we aim at choosing Δ to minimize $R(\theta, P_{\text{te}}(\mathbf{x}, w))$.

Note that even if the alignment is done appropriately, the joint distributions between training and testing could differ because $P_{\text{tr}}(\mathbf{x}) \neq P_{\text{te}}(\mathbf{x})$ due to a form of sample bias. A proper alignment simplifies the data set shift problem into a covariate shift problem.

B. A Maximum Likelihood Approach.

Our methodology to attack the problem above consists of shifting T_{te} using maximum likelihood. To begin, we first assume the marginal distribution from which the training is drawn, $P_{\text{tr}}(\mathbf{x})$, is a mixture of Gaussians. We estimate parameters directly from our sample T_{tr}, since we know all class labels (i.e., we know which vector belong to each component or Gaussian). This enables us to have a complete characterization of the marginal distribution:

$$P_{\text{tr}}(\mathbf{x}) = \sum_{i=1}^{c} \phi_i \; g_i(\mathbf{x}|\mu_i, \Sigma_i) \tag{2}$$

$$g_i(\mathbf{x}|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i)\} \tag{3}$$

where ϕ_i, μ_i, and Σ_i are the mixture coefficient (i.e., prior probability), mean and covariance matrix of the ith component respectively, n is the number of features, and c is the number of components.

Our next step is to define a new testing set $T'_{\text{te}} = \{\mathbf{x}'\}$, where $\mathbf{x}' = (x_1 + \delta_1, x_2 + \delta_2 + ... + x_n + \delta_n)$, since we know a shift has occurred along our input features. Our approach is to find the Δ that maximizes the log likelihood of T'_{te} with respect to distribution $P_{\text{tr}}(\mathbf{x})$:

$$\mathcal{L}(\Delta|T'_{\text{te}}) = \log \prod_{k=1}^{q} P_{\text{tr}}(\mathbf{x}^k) = \sum_{k=1}^{q} \log P_{\text{tr}}(\mathbf{x}^k) \tag{4}$$

To simplify our notation, we can rewrite the marginal distribution (equation 2) as follows:

$$P_{\text{tr}}(\mathbf{x}^k) = \sum_{i=1}^{c} \alpha_i v_i(\mathbf{x}^k) \tag{5}$$

where

$$\alpha_i = \phi_i \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}}$$

and

$$v_i(\mathbf{x}^k) = \exp\{-\frac{1}{2}(\mathbf{x}^k - \mu_i)^T \Sigma_i^{-1}(\mathbf{x}^k - \mu_i)\}$$

When we apply a shift across the set of features, the resultant marginal test distribution can be modelled as follows:

$$P_{\text{te}}(\mathbf{x}^k) = \sum_{i=1}^{c} \alpha_i z_i(\mathbf{x}^k) \tag{6}$$

where

$$z_i(\mathbf{x}^k) = \exp\{-\frac{1}{2}(\mathbf{x}^k + \delta - \mu_i)^T \Sigma_i^{-1}(\mathbf{x}^k + \delta - \mu_i)\}$$

Here we define a new vector δ that captures the amount of shift applied to each feature; vector δ has values other than zero only for the set of features in X_B (Section III-A).

The equation for the log-likelihood function ends up as follows:

$$\mathcal{L}(\Delta|T'_{\text{te}}) = \sum_{k=1}^{q} \log \left(\sum_{i=1}^{c} \alpha_i z_i(\mathbf{x}^k) \right) \tag{7}$$

To solve this optimization problem, we use an iterative gradient ascent approach; we search the space of values in Δ for which the log-likelihood function reaches a maximum value. We stop the iteration when the change in the log-likelihood function is less than a small value ϵ (in our experiments $\epsilon = 0.001$).

IV. Related Work

Domain adaptation has gained much attention recently, mainly due to the pervasive character of problems where distributions change over time; it assumes the learning task remains constant (e.g., classifying variable Cepheid stars into two classes), but the marginal and class posterior distributions between source and target domain may differ (as opposed to transfer learning, where tasks can exhibit different input representations. i.e., different input spaces).

Domain adaptation has been attacked from different angles: by searching for a single representation that unifies both source and target domains [11], [12], [13]; by proving error bounds as a function of empirical error and the distance between source and target distributions [1], [2], [14]; as part of multitask learning by finding a common representation using spectral functions [15]; within a co-training framework where target vectors are incorporated into the source training set based on confidence [16]; by re-weighting source instances [17]; and by using regularization terms to learn models that perform well on both source and target domains [18]. Different from previous work, our strategy is to shift the target data to achieve a close alignment with the source data. By making both sets as similar as possible, we expect the source model to generalize well on the target domain.

V. Light Curve Feature Characterization.

Most previous work in lightcurve classification deals with features useful over a broad number of classes. As an example, identification of microlensing events requires features that can discriminate among several variable sources such as periodic variables, eruptive variables, and supernovae [19], [20]. Typically, features are extracted from spectral analysis, and can be specialized or generic; the latter includes statistical tests, coefficients from the Fourier transform, and autocorrelation coefficients. In contrast, our study targets specialized geometric properties that serve as better discriminators between the two types of pulsation mode in Cepheid variable stars (fundamental and first-overtone). We employ five features out of which three -Phase Shift, Depth, and Slope- are calculated by fitting a polynomial curve along the points of one pulsation cycle. The other two features, Mean Magnitude and Log Period, are directly available from the Optical Gravitational Lensing experiment (Section VI). Figure 2 shows our features along typical light curves for the two different pulsation modes. Phase Shift is the phase at which the light curve shows its highest magnitude (apparent brightness) in one phase cycle; Depth is equal to the difference between the highest and lowest magnitudes in one complete pulsation cycle; Slope is the angle of the line that connects the highest and lowest magnitudes. These features point to stark differences in the geometric shape of light curves between the two classes of interest, and thus appear more informative than other more typical features.

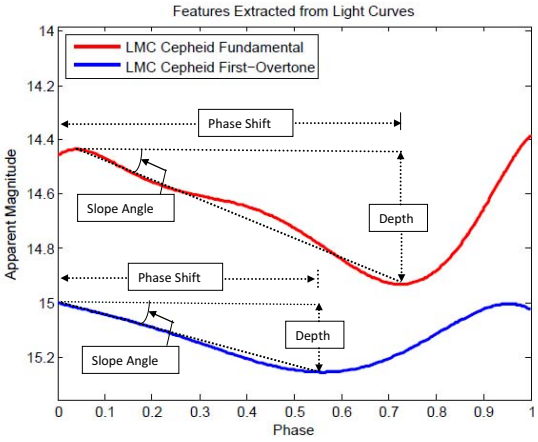

Fig. 2. Features capture geometric properties of light curves that exhibit clear differences between the two classes of interest.

VI. Experiments

Our experiments assume the Large Magellanic Cloud (LMC) galaxy as the source domain, and M33 and Small Magellanic Cloud (SMC) galaxies as target domains. We use data from the third phase of the Optical Gravitational Lensing Experiment [21] (OGLE-III)[2] and the M33 Synoptic Stellar Survey[3]. We analyze the third catalog of OGLE Large Magellanic Cloud Cepheids [22]; for space considerations we report on the infrared band only; the training set contains 3,056 points (i.e., light curves). The M33 Synoptic Stellar Survey [23] is a follow-up study of Cepheids and other variables initially discovered by the DIRECT project [24], [25]; the testing set contains 322 points (infrared band). The Small Magellanic Cloud data is also available at the OGLE repository and contains 4,254 points (infrared band).

As a pre-processing step, parameters for the mixture of Gaussians on the source domain are estimated directly through the data. We assume two Gaussians (one Gaussian for class fundamental and one for overtone). Since class labels are available, it is straightforward to estimate means, covariance matrices and mixing proportions. Implementation of the learning algorithms can be found in WEKA [26] using default parameters. Specifically, neural networks are invoked with

[2]Data repository available at the following address: http://ogledb.astrouw.edu.pl/ ogle/CVS/

[3]Data repository available at the following address: http://faculty.physics.tamu.edu/lmacri/M33SSS/

TABLE I. CLASSIFICATION ACCURACY WITH AND WITHOUT DATA ALIGNMENT.

Learning Algorithm	LMC model tested on SMC			LMC model tested on M33		
	No Data Alignment	With Data Alignment		No Data Alignment	With Data Alignment	
		Using magnitude	Using all features		Using magnitude	Using all features
Neural Networks	91.31 (0.29)	98.03 (0.11)	99.41 (0.17)	91.82 (1.42)	91.82 (1.42)	92.76 (1.09)
SVM (PolyKernel 1)	98.01 (0.19)	98.67 (0.19)	98.67 (0.19)	91.82 (1.42)	91.82 (1.42)	95.21 (1.00)
SVM (PolyKernel 2)	97.42 (0.10)	98.88 (0.17)	98.96 (0.19)	91.82 (1.42)	91.82 (1.42)	95.42 (0.89)
SVM (PolyKernel 3)	97.95 (0.10)	98.92 (0.18)	99.18 (0.18)	91.82 (1.42)	91.82 (1.42)	94.91 (0.75)
J48	96.53 (0.26)	96.58 (0.27)	98.90 (0.19)	91.64 (1.60)	91.64 (1.60)	93.79 (1.00)
Random Forest	98.23 (0.22)	98.34 (0.22)	99.21 (0.25)	91.94 (1.42)	91.94 (1.42)	94.24 (0.89)

one hidden layer, two internal nodes, a learning rate of 0.3, momentum of 0.2, and 500 epochs. Support Vector Machines are invoked with polynomial kernels of degrees 1, 2, and 3 respectively. Decision trees are invoked with a confidence factor of 0.25. Random Forests combine 10 decision trees on each run.

Our experiments compute predictive accuracy when the model obtained on LMC (source domain) is applied to the transformed target sample (where a shift Δ is added to each feature vector, Section III-B). Accuracy is obtained by creating bootstrap samples of the target or test set (SMC and M33); each sample is generated by sampling with replacement. We generated 10 samples, and averaged our results after applying the training model on each transformed sample. Table 1 shows our results. The first column corresponds to the learning algorithms. The second column shows accuracy values on SMC with no data alignment. The third and fourth columns show accuracy on SMC with data alignment; we make a distinction between shifting along apparent magnitude only vs shifting over all features[4] (including our geometric features, section V). Columns 5-7 are equivalent to columns 2-4, but with M33 as the target domain. Numbers in parentheses represent standard deviations.

Our results show a significant increase in accuracy when the data alignment step is applied to the target sample, and all features are used during data alignment ($p = 0.05$ level using a t-student distribution); this is true for all learning algorithms under consideration. There is also an advantage when data alignment is effected over all features, as opposed to shifting along magnitude only, but the differences are not always significant. The mean increase in accuracy over all learning algorithms with SMC as the target domain is 1.66% when shifting along magnitude, and 2.48% when shifting along all features. Equivalent results with M33 as the target domain show no increase in accuracy when shifting along magnitude, and 2.57% when shifting along all features. Overall our results show the effectiveness of the data alignment step, and the geometric feature representation.

Additional observations can be drawn from Table 1. For example, accuracy gain is more clear when using Neural Networks, Decision Trees, and SVMs with high order polynomials; these algorithms tend to generate complex decision boundaries, which suggests a relatively complex class distribution. In addition, it is interesting to observe that the gain in accuracy between LMC and M33 is evident when shifting along all features, but there is no accuracy gain when shifting along magnitude. A deeper analysis reveals that, even after aligning the source and target datasets, some samples show high dispersion, which complicates any attempt to re-use the model built on the source domain. To gain a better understanding of this problem, we measured the distance between the data distributions of LMC and M33. Our study relies on the Kullback-Leibler (K-L) distance, also called relative entropy, defined as follows:

$$D_{\mathrm{KL}}(P||Q) = \int_{-\infty}^{\infty} \ln \frac{p(x)}{q(x)} p(x) dx \tag{8}$$

where P and Q are the two distributions whose distance we intend to quantify, and $p(x)$ and $q(x)$ are the probability densities of P and Q respectively. We used three techniques to approximate K-L distance [27]. Table 2 shows our results. The first column shows the approximation techniques to compute K-L distance. The second column shows the degree of divergence of LMC vs M33 with no data alignment, the third and fourth columns show the corresponding divergence of LMC vs M33 when shifting along magnitude and sifting along all features respectively. There is a significant reduction in K-L distance when using the data alignment step, but results in Table 2 do not correlate clearly with accuracy gain. The increase in accuracy on M33 between shifting along magnitude vs shifting along all features is of 2.57%, but there is no increase in accuracy between no data alignment and shifting along magnitude, which corresponds to the largest reduction in K-L distance. We attribute this to a phase transition effect, where crossing a threshold of data similarity between the source and target data, is needed to trigger a boost in accuracy increase. The effect is a sharp spike in accuracy, rather than a gradual increase.

VII. SUMMARY AND CONCLUSIONS

We propose an approach to domain adaptation where a model trained on a source domain can be re-used on a target domain. Our study revolves around the automated prediction of Cepheid stars from different galaxies, where each star is characterized using geometrical properties of its corresponding light curve. Specifically, we propose a framework where the model learned over data from a source galaxy can be adapted to data from a target galaxy. Our methodology is based on using maximum likelihood based on a mixture of Gaussian models; the goal is to shift the target data towards the source data. Our experiments show an increase in performance accuracy when we invoke the data alignment step, and when we incorporate our geometric feature representation.

Future work will explore more closely the connection between the degree of distribution divergence, and the effectiveness of the data alignment step proposed in Section III-B.

[4]When shifting along all features we exclude the log period, as this is independent of the distance to galaxies.

TABLE II. KULLBACK-LIEBLER DIVERGENCE BEFORE AND AFTER DATA ALIGNMENT.

K-L Divergence	LMC vs M33		
	No data alignment	Using magnitude	Using all features
Monte Carlo	257.28 (0.78)	7.05 (0.24)	4.76 (0.24)
Matching Based	123.02 (25.44)	4.83 (1.24)	3.49 (0.66)
Unscented Transform	128.93 (1.68)	3.94 (1.14)	2.42 (0.71)

The goal will be to find an automated mechanism to indicate when the re-use of a training model obtained in a source domain can be effectively applied to a new target domain. We also plan to extend our ideas to the context of transfer learning [28]. Dataset shift resembles transfer learning, but while the former deals with changes on the same task, the latter comprises a broader setting where target tasks can differ substantially from the source task [3]. We plan to explore when a model can be re-used, based on the presence of invariant global data properties across tasks.

ACKNOWLEDGMENTS

We acknowledge support for this project by NASA, Space Telescope Science Institute, the National Science Foundation, National Optical Astronomy Observatory, Texas A&M University, and by the University of Houston.

REFERENCES

[1] S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, Analysis of Representations for Domain Adaptation, NIPS (2006).

[2] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J. Wortman, A Theory of Learning from Different Domains, Machine Learning, Special Issue on Learning From Multiple Sources, 79 (2010) 151–175.

[3] A. Storkey, When training and test sets are different, in: J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, N. D. Lawrence (Eds.), Dataset Shift in Machine Learning, MIT Press, 2009, pp. 3–28.

[4] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, N. D. Lawrence, Dataset Shift in Machine Learning, MIT Press, 2009.

[5] H. Shimodaira, Improving predictive inference under covariate shift by weighting the log-likelihood function, Journal of Statistical Planning and Inference 90 (2000) 227–244.

[6] M. Sugiyama, N. Rubens, K. R. Muller, A conditional expectation approach to model selection and active learning under covariate shift, in: J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, N. D. Lawrence (Eds.), Dataset Shift in Machine Learning, MIT Press, 2009, pp. 107–130.

[7] J. P. Cox, Theory of Stellar Pulsation, Princeton University Press, 1980.

[8] J. W. Richards, D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy, M. Rischard, H. Shimodaira, On machine-learned classification of variable stars with sparse and noisy time-series data, The Astrophysical Journal 733:10.

[9] J. Debosscher, L. M. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido, E. Solano, Automated supervised classification of variable stars, Astronomy & Astrophysics 475 (2007) 1159–1183.

[10] R. Vilalta, K. Dhar Gupta, L. Macri, A Machine Learning Approach to Cepheid Variable Star Classification using Data Alignment and Maximum Likelihood, Astronomy and Computing 2 (2013) 46–53.

[11] J. Blitzer, R. McDonald, F. Pereira, Domain Adaptation with Structural Correspondence Learning, Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics(2006).

[12] R. Ando, T. Zhang, A framework for learning predictive structures from multiple tasks and unlabeled data, Journal of Machine Learning Research JMLR, 6, 2005, pp. 1817-1853.

[13] X. Glorot, A. Bordes, Y. Bengio, Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach, International Conference on Machine Learning (2011). Astronomy & Astrophysics 475 (2007) 1159–1183.

[14] J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J. Wortman, Learning Bounds for Domain Adaptation, Neural Information Processing Systems NIPS (2007).

[15] A. Argyriou, C. Micchelli, M. Pontil, Y. Ying, A Spectral Regularization Framework for Multi-Task Structure Learning, Neural Information Processing Systems NIPS (2007).

[16] M Chen, K, Q. Weinberger, J. Blitzer, Co-Training for Domain Adaptation, Neural Information Processing Systems NIPS (2011).

[17] Y. Mansour, M. Mohri, A. Rostamizadeh, Domain Adaptation with Multiple Sources, Advances in Neural Information Processing Systems, NIPS (2009) pp. 1041–1048.

[18] H. Daume, A. Kumar, A. Saha, Co-regularization Based Semi-supervised Domain Adaptation, Neural Information Processing Systems NIPS (2010).

[19] 5. V. Belokurov, N. W. Evans, Y. Le Du, Light-curve classification in massive variability surveys I. Microlensing., Mon. Not. R. Astron. Soc. MNRAS, 341 (2003) 1373–1384.

[20] 5. V. Belokurov, N. W. Evans, Y. Le Du, Light-curve classification in massive variability surveys II. Transients towards the Large Magellanic Cloud., Mon. Not. R. Astron. Soc. MNRAS, 352(1) (2004) 233–242.

[21] A. Udalski, The optical gravitational lensing experiment (ogle): Bohdan's and our great adventure, in: K. Z. Stanek (Ed.), The Variable Universe: A Celebration of Bohdan Paczynski, Vol. 403 of Astronomical Society of the Pacific Conference Series, 2009, p. 110.

[22] I. Soszynski, R. Poleski, A. Udalski, M. K. Szymanski, M. Kubiak, G. Pietrzynski, L. Wyrzykowski, O. Szewczyk, K. Ulaczyk, The optical gravitational lensing experiment. the ogle-iii catalog of variable stars. i. classical cepheids in the large magellanic cloud, Acta Astronomica 58 (2008) 163–185.

[23] A. Pellerin, L. M. Macri, The m33 synoptic stellar survey. i. cepheid variables, Astrophysical Journal Supplement Series 193 (2011) 26.

[24] K. Z. Stanek, J. Kaluzny, M. Krockenberger, D. D. Sasselov, J. L. Tonry, M. Mateo, Direct distances to nearby galaxies using detached eclipsing binaries and cepheids. ii. variables in the field m31a, Astronomical Journal 115 (1998) 1894–1915.

[25] L. M. Macri, K. Z. Stanek, D. D. Sasselov, M. Krockenberger, J. Kaluzny, Direct distances to nearby galaxies using detached eclipsing binaries and cepheids. vi. variables in the central part of m33, Astronomical Journal 121 (2001) 870–890.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The weka data mining software: An update, SIGKDD Explorations 11 (1).

[27] J. Goldberger, S. Gordon, H. Greenspan, An Efficient Image Similarity Measure Based on Approximations of K-L Divergence Between Two Gaussian Mixtures, Proceedings of the Ninth IEEE International Conference on Computer Vision (2003).

[28] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345–1359.