A Novel Metric Combination Approach for Verification Problems

© Archil Maysuradze
Lomonosov Moscow State University,
Moscow, Russia
maysuradze@cs.msu.ru
© Bharathi Pilar
University College,
Mangalore, India
bharathi.pilar@gmail.com
© B.H. Shekar
Mangalore University,
Mangalore, India
bhshekar@gmail.com

Abstract. In this paper we propose and test a novel approach, namely Supervised Asymmetric Metric Extraction (SAME), that learns from the supervised metric data and extracts the best single metric from a given set of metrics. It takes up large space to represent the metric-based descriptions, so the approach is specifically crafted to allow for a computationally effective solution. The proposed learning model is scale-independent and hence rescaling of any metric does not affect the learning. Another advantage of metric extraction is the way of training set annotation which specifically suits verification problems. In this metric extraction approach, we separate intraclass and interclass distances, simplifying the metric extraction problem to linear programming problem which can use optimization techniques effectively. Here, the number of variables needed in the computation remains small and hence it eliminates the need of any soft constraint extension resulting in reduced computational time complexity. The experimental results on offline and online signature data demonstrate that the proposed approach yields better performance and time complexity compared to other metric extraction technique as computational complexity in the proposed approach depends mainly on the calculation of original distances.

Keywords: verification problem, problem-specific annotation, multiple distance metrics, metric-based descriptions, dimensionality reduction, metric combination, constrained optimization.

1 Introduction

Due to the rapid development in electronics, communication and hardware technology, there is a high demand for the design of automated intelligent systems in industrial works, medical imaging, defense, data analytics and biometrics. In machine learning theory and data mining applications, there are increasingly frequent situations when different ways to measure similarity are set on the same objects. Such situations are typical in information retrieval, computer vision, biology, social systems, finance, etc. In many of these domains, similarity engineering is an important way to incorporate expert knowledge and similarity learning is a way to produce a similarity function based on some constraints. The performance of such automated intelligent systems depends upon suitable choice of similarity/dissimilarity function over the input space. Hence there is a increasing need of an extraction mechanism which can learn a best metric for a given set of data from the given set of several metrics. In this context, in this work we have presented metric extraction technique wherein the actual distance learning is performed only on a finite set of supervised training data, whereas the best metric learnt can be applied to the whole population of objects.

The conventional machine learning approaches focus on feature extraction and feature learning, here we have adopted a novel approach of metric extraction from several metrics. The idea here is, even if the individual metrics fail to discriminate classes accurately, their combination will definitely improve the quality of the discrimination. Hence, there is a need for aggregating original multimetric information. In this context, in this work we recognize a notion of metric dimension reduction methods which produces a best metric for a given population from a set of original metrics applied on a small set of labelled samples from the population. The theoretical properties of the problem are provided along with experimental results to exhibit the performance of the proposed approach. The computational complexity will be low as it is mainly due to calculation of original distances.

Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018

In automated intelligent systems, specifically biometrics, there are mainly two types of classification problems. First, identification systems that recognizes the subject by its description. Second, verification systems that test whether the given description relates to the given subject. In this paper, we focus on the verification systems, namely biometric systems that verify a person by his/her signatures.

In certain applications like mobile banking, as we need to deal with an intensive stream of verification queries, the update of the system with new information must be simple. This is met by means of specific annotation and classification techniques.

2 Related Work

In the state of art literature there are ample number of distance metric learning approaches applied to machine learning problems such as computer vision, biometrics, information retrieval and data analytics [2,7]. All these methods can be broadly categorized into two main categories, one is eigenvalue optimization and the other is convex or non-convex optimization. Most of the popular methods are based on the ideas of Large-Margin nearest neighbor [14] and Information-theoretic metric learning [5]. The formulations of metric learning are similar to multiple kernel learning which is very popular in the field of machine learning. It is theoretically proved in [10,11] that the metric investigation approaches are computationally effective when the derived metric is a linear combination of the original ones compared to nonnegative linear or convex combinations of original metric.

From the literature we can notice, the conventional machine learning algorithms are based on feature learning and metric learning. The metric learning produces new metric from conventional feature-based object descriptions. Our paper focuses on metric extraction technique which is a kind of dimensionality reduction of metric-based descriptions. The approach does not focus on object features but only takes specific annotations for pairs of objects. The conventional machine learning approaches ensembles the different classifiers at feature level or decision level, thus combining several distances obtained by various metrics in order to have consensus decision, whereas distance extraction technique aggregates original multimetric information giving rise to the best metric. In addition, most of the conventional machine learning approaches use non-negative linear or convex combinations of multiple metrics making it computationally expensive, whereas our approach is computationally efficient as the derived metric is a linear combination of original metric with a scalar value.

3 Proposed Approach

3.1 Distance combination model

We consider a multimetric object space with N pseudo-metrics. Let ρ_1, ρ_2, ..., ρ_N be the original metrics on this object space. Let r be the new derived metric which is linear combination of original metrics defined as follows,

$$r(x, y) = w_1\rho_1(x, y) + w_2\rho_2(x, y) + \cdots + w_N\rho_N(x, y)$$

where w_1, w_2,..., w_N are the weights. The derived distance $r(x, y)$ can be calculated for any pair (x, y) from the whole population. The combination model is rather popular in the literature [6].

It is guaranteed that r is a pseudo-metric if the weights are non-negative [11]. The guarantee is important for theoretical correctness of many metric methods of machine learning and artificial intelligence.

3.2 Annotation model

Let T be the finite sample of size M, and x_1, x_2,..., x_M be the objects of the sample. There are two sets of unordered pairs from the sample of training objects: the first set is annotated as must-link pairs, the other as cannot-link pairs. In the conventional case when each training object is annotated with its class label, the must-link set contains all the pairs of the same class training objects and the cannot-link set contains all the pairs of training objects from different classes.

But for verification problems we propose to use another approach to annotation. In case of signature verification systems, we expect that for each person there will be genuine signatures and skilled forgeries. Accordingly, must-link pairs will represent pairs of genuine signatures, whereas cannot-link pairs will represent pairs of genuine signatures and skilled forgeries. Notice that we do not include the pairs of skilled forgeries to the training at all.

This annotation model is more expressive than labelling individual objects. Moreover, when we will use the extracted (derived) metric for verification we will have several options to prototype selection. One of the crucial advantages is that we will be able to add new people to the verification system without prototyping any forgeries.

One of the important challenges for training an automated signature verification system is the presence of partial knowledge during training. In a realistic scenario, during training we only have access to genuine signatures for the users enrolled to the system. During operations, however, we want the system not only to be able to accept genuine signatures, but also to reject forgeries. This is a challenging task, since during training a classifier has no information to learn what exactly distinguishes a genuine signature and a forgery for the users enrolled in the system. The proposed approach directly addresses this problem as our approach just takes only must link pair between samples of genuine class. We do not need forgery class.

3.3 Recognition model

As mentioned above, recognition process must deal with people without known forgeries. For that reason, we utilize the property of scale-independence which we are going to discuss in the following section. Our learning method automatically rescales the derived metric, so we can calculate the distances to genuine prototypes only. Then the distance to the nearest genuine prototype may be used as a score for classification.

We need forgeries for metric extraction, but later we can add a new person to our verification system without any forgery samples.

Conventional performance metrics for verification problems are FAR (false acceptance rate) and FRR (false rejection rate). As our recognition model produces a score, we can use ROC-analysis and calculate EER (equal error rate) which is an informative integral performance metric for the problem in hand.

At recognition phase, all the original metrics to all the prototypes can be calculated in parallel, so the decision can be made quickly.

3.4 Learning model

Let S be the set of must-link pairs, let D be the set of cannot-link pairs. Here, the reflexive pairs (i, i) are not considered. Let $y_1, \ldots, y_M$ be the actual labels (Genuine or Forgery), then we can set $S = \{(i, j) \mid y_i = y_j = G\}$ and $D = \{(i, j) \mid y_i = G \text{ and } y_j = F\}$, where S and D are intraclass and interclass pairs respectively. The metric extraction problem here is to minimize the average intraclass derived distance provided all interclass derived distances are not less than 1 and the weights $w_1, w_2, \ldots, w_N$ are non-negative. The problem is formalized as follows

$$\sum_{(i,j)\in S} r(x_i, x_j) \to \min$$

$$\text{s.t.} \quad r(x_i, x_j) \geq 1, \text{for } (i, j) \in D$$

$$w_n \geq 0, \text{for } n \in \{1, \ldots, N\}.$$

When we use the linear form of the derived metric r and change the order of summation in the objective function, we get the conventional linear programming problem

$$\sum_{n=1}^{N} w_n \sum_{(i,j)\in S} \rho_n(x_i, x_j) \to \min$$

$$\text{s.t.} \quad \sum_{n=1}^{N} w_n \rho_n(x_i, x_j) \geq 1, \text{for } (i, j) \in D \qquad (1)$$

$$w_n \geq 0, \text{for } n \in \{1, \ldots, N\}.$$

The formalization makes it possible to consider the result as metric selection by positive weights. As the scales are ignored in the approach, so the weight rank has no semantics. The approach selects relevant metrics. To get rid of redundant metrics, it is advised to use unsupervised metric extraction in advance [9]. This is a linear programming problem, hence there is a wide range of methods and software to find the global optimum. In fact, we solve the dual problem to (1).

Definition 1. A pair of objects from D is conflicting, if for all the original metrics the distance between the objects is zero.

Theorem 1. In the linear programming problem, there is always an admissible solution, unless there are conflicting objects in the sample.

It follows from the theorem that we can do without slack variables customary for SVM. A soft-margin extension is not required at all. Consequently, the number of variables and the size of the optimization problem remains small which is crucial to maintain the computational complexity at low levels.

3.5 Scale-independence

The important feature of the proposed learning model is its scale-independence. This means that any rescaling of any original metric does not affect the learning. This is unlike many conventional feature-based techniques where changing the scale or units of a feature - e.g. from meters to kilometers – may change the results dramatically. Also, any derived metric will have a standard scaling for the threshold from the learning model. That is why distances produced by any derived metric may be used as coherent scores for classification. Hence, we do not need samples of forgeries for recognition.

4 Applications

Let us consider automatic online signature verification problem in Biometric domain which is used to prevent identity fraud by verifying the authenticity of signatures. The dataset consists of both genuine and forged signatures. Although, there are plethora of metrics for computer vision applications such as online signature verification, devising an efficient and accurate metric is still a challenging issue. Hence, we are motivated to develop a computationally efficient metric extraction approach that can find linear combinations of the original metrics such that the resulting classification error is reduced compared to that of the original metric.

5 Datasets and Original Metrics

5.1 Offline signature verification

For the offline signature verification problem, we used Centre of Excellence for Document Analysis and Recognition (CEDAR) dataset. The CEDAR at SUNY Bualo has built the offline signature dataset with 55 signers, a total of 2640 signature samples. 24 genuine signature samples were collected from each signer and later, to obtain the forgeries (skilled), 20 arbitrary chosen signers skilfully forged the signature in the dataset each with 24 samples. Hence for each signer, 24 genuine and 24 skilled forge samples, a total of 48 signature samples were collected. Each signature image was labelled, i.e. it was known which person it had been taken from and whether it is genuine or forged.

In our experiment, we have partitioned each signature image into eight vertical partitions of equal width. The morphological pattern spectrum-based features are extracted from the signature partitions as explained in our earlier paper [13]. The dominant features of each partition in a signature image is represented in the form of histogram. There are eight histograms corresponding to 8 partitions of each signature. All these feature vectors are stored in the dataset. The Earth Mover's Distance (EMD) metric is used to compare the histograms. This produced a set of eight metrics, each comparing characteristics of an individual and small part of the whole image.

5.2 Online signature verification

For the online signature verification problem, we used two datasets: SVC2004 [15] and a rather new MOBISIG [1]. The use-case scenario for MOBISIG may be quite data intensive as it assumes that many people will be querying the verification system at the same time and fraudsters may attack the system with many images. Again, the samples were divided into training set and test set.

We used 8 original distance functions. All of them are known to have been state-of-the-art for the problem when used individually.

15. DTW for (x; y) – DTWxy [4];
16. DTW for (vx; vy) – DTWv [4];
17. ER2 for (x; y) – ER2xy [8];
18. ER2 for (vx; vy) – ER2v [8];
19. EMD for angles – EMDth [12];
20. EMD for vx – EMDvx [12];
21. EMD for vy – EMDvy [12];
22. SumMinxy(x; y)

6 Experimental Results

The Table 1 and Table 2 below show the equal error rates for each of the original metrics and for the learned metric for offline and online signature verification problems respectively.

The Table 1 shows EER for 8 individual original metrics and derived metrics for the offline dataset. EER is averaged over people. As one can expect, none of the original metrics can be good for classification separately, as each one contains only one eighth of information about the signature. At the same time, the combination of these metrics can result in a considerably good discriminative function.

Table 1 The equal error rates for offline signature verification problem (%)

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	r
84.2	89.7	86.2	89.5	86.6	84.8	88.7	85.0	23.0

The Table 2 shows EER for 8 individual original metrics and derived metrics for both online datasets. EER is averaged over people. MMC [3,7] is the most cited metric learning technique. We converted it to be a metric extraction technique as a competitor.

Table 2 The equal error rates for online signature verification problem (%)

Metric	SVC2004	MOBISIG
DTWxy	8.00	7.23
DTWv	10.00	8.43
ER2xy	8.75	8.67
ER2v	15.00	9.88
EMDth	16.75	14.94
EMDvx	20.50	25.06
EMDvy	26.25	22.65
SumMinxy	12.50	13.49
MMC	9.00	8.43
SAME	**5.00**	**4.80**

The metric extraction phase could be done once for a set of people. Metric combination and verification are extremely quick - in fact, it is negligible in comparison with image or time series transfer and preprocessing.

7 Conclusion

In this paper we presented a novel metric extraction method which produces best metric for a given population from a linear combination of several original metrics applied on a finite set of supervised training samples. In this approach the intraclass and interclass distances are treated separately resulting in linear optimization problem with reduced computational time complexity. The experimental results on offline signature as well as online Signature verification demonstrated the significant reduction in the error rate compared to the individual metrics. In addition, it is also shown that the metric extraction procedure is computationally effective taking only seconds to calculate coefficients of the linear combination.

Acknowledgments. This research is partially supported by Russian Foundation for Basic Research (projects 16-01-00196 and 17-01-00917) and Lomonosov Moscow State University (project 5.1.16).

References

[1] Antal, M., Szabo, L. Z., Tordai, T.: Online Signature Verification on MOBISIG Finger-Drawn Signature Corpus. In: Mobile Information Systems (2018) doi: 10.1155/2018/3127042

[2] Bellet, A., Habrard, A., Sebban, M.: A Survey on Metric Learning for Feature Vectors and Structured Data. CoRR abs/1306.6709 (2013)

[3] Bellet, A., Habrard, A., Sebban, M.: Metric Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, San Rafael (2015)

[4] Berndt, D. J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: KDD workshop, V. 10, N. 16, pp. 359-370 (1994)

[5] Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th international conference on Machine learning, pp. 209–216. ACM (2007) doi: 10.1145/1273496.1273523

[6] Jang, D., Jang, S.-J., Lim, T.-B.: Distance combination for content identification system. In: 1st International Conference on Communications, Signal Processing, and their

Applications (ICCSPA), pp. 1–6 (2013) doi: 10.1109/ICCSPA.2013.6487261

[7] Kulis, B.: Metric learning: a survey. In: Foundations and Trends in Machine Learning, vol. 5, no. 4, pp. 287–364 (2013). doi: 10.1561/2200000019

[8] Lei, H., Palla, S., Govindaraju, V.: ER/sup 2: an intuitive similarity measure for on-line signature verification. In: Frontiers in Handwriting Recognition, pp. 191-195. IEEE (2004) doi: 10.1109/IWFHR.2004.38

[9] Maysuradze, A.I., Suvorov, M.A.: Aggregation of multiple metric descriptions from distances between unlabeled objects. In: J. Comput. Math. Math. Phys. 57(2), pp. 350–361 (2017) doi: 10.1134/S0965542517020105

[10] Maysuradze, A.I.: Homogeneous and rank bases in spaces of metric configurations. In: J. Comput. Math. Math. Phys. 46(2), pp. 330–344 (2006) doi: 10.1134/S096554250602014X

[11] Maysuradze, A.I.: On optimal decompositions of finite metric configurations in pattern recognition problems. In: J. Comput. Math. Math. Phys. 44(9), pp. 1615–1624 (2004)

[12] Rubner, Y., Tomasi, C., Guibas, L. J.: A metric for distributions with applications to image databases. In: Computer Vision, pp. 59-66. IEEE (1998) doi: 10.1109/ICCV.1998.710701

[13] Shekar, B. H., R. K. Bharathi, Bharathi Pilar: Local morphological pattern spectrum based approach for off-line signature verification. In: International Conference on Pattern Recognition and Machine Intelligence. Springer, Berlin, Heidelberg (2013) doi: 10.1007/978-3-642-45062-4_45

[14] Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: J. Mach. Learn. Res. 10, pp. 207–244 (2009)

[15] Yeung, D. Y. et al.: SVC2004: First international signature verification competition. In: Biometric Authentication, pp 16-22. Springer, Berlin, Heidelberg (2004)