GEOMETRICAL TOOLS FOR ALPHA-VORONOI PARTITIONS

Atsumi Ohara and Yuya Nagatani

Abstract: *We consider problems to classify data represented as discrete probability distributions. For the classification we propose the Voronoi partition technique with respect to α-divergence, which is a statistically justified pseudo-distance on the space of probability distributions. In order to improve computational efficiency and performance of the classification, we introduce two nonlinear transformations respectively called the escort and projective transform, and weighted α-centroids. Finally we demonstrate performances of the proposed tools via simple numerical examples.*

Keywords: *α-divergence, α-Voronoi partition, Escort transformation, Projective transformation*

ACM Classification Keywords: *I.3.5 Computational Geometry and Object Modeling*

MSC: *53A15, 68T10*

Introduction

The Voronoi partitions on the space of probability distributions with the Kullback-Leibler or Bregman divergences have been recognized as significant approaches for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on [1, 2, 3, 4].

Recently we have proposed [7] (See [8] for the proofs and derivations) a computationally efficient method to construct Voronoi diagrams with respect to the α-divergence [5, 6] using conformal flattening technique. One of mathematically natural requirements for divergence functions is an *invariance* property and it is proved that the α-divergence is equipped with the property [5, 6]. Hence the α-divergence is an important candidate to be applied to the above statistical problems.

In this note, we introduce several useful tools when we apply the α-Voronoi partitions on the space of discrete probability distributions for the purpose of pattern classification problems.

Preliminaries

Let $\mathcal{S}^n$ denote the n-dimensional probability simplex, i.e.,

$$\mathcal{S}^n := \left\{ \boldsymbol{p} = (p_i) \;\middle|\; p_i > 0,\ \sum_{i=1}^{n+1} p_i = 1 \right\}, \tag{1}$$

and $p_i, i = 1, \cdots, n+1$ denote probabilities of $n+1$ states.

The α-divergence [5, 6] is a function on $\mathcal{S}^n \times \mathcal{S}^n$ defined for $\alpha \neq \pm 1$ by

$$D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = \frac{4}{1-\alpha^2} \left\{ 1 - \sum_{i=1}^{n+1} (p_i)^{(1-\alpha)/2} (r_i)^{(1+\alpha)/2} \right\}.$$

Note that it respectively converges to the Kullback-Leibler divergence or its dual when α goes to -1 or 1, and $D^{(0)}$ is called the Hellinger distance.

For $q \in \mathbf{R}$ escort transformation [9] is defined for $\boldsymbol{p} \in \mathcal{S}^n$ by

$$P_i(\boldsymbol{p}) := \frac{(p_i)^q}{\sum_{j=1}^{n+1}(p_j)^q}, \; i = 1, \cdots, n+1, \quad Z_q(\boldsymbol{p}) := \sum_{i=1}^{n+1} \frac{(p_i)^q}{q}, \tag{2}$$

and we call $P_i(\boldsymbol{p})$ an escort probability. Note that the escort distribution $\boldsymbol{P}(\boldsymbol{p}) = (P_i(\boldsymbol{p}))$ converges to the uniform distribution independently of $\boldsymbol{p}$, when $q \to 0$. On the other hand, when $q \to \pm\infty$, it converges to a distribution on the boundary of $\mathcal{S}^n$ depending on the maximum or minimum components of $\boldsymbol{p}$. In the sequel we fix the relation between α and q by $q = (1+\alpha)/2$, and assume $q > 0$.

The conformal divergence [7] for $D^{(\alpha)}$ is defined by

$$\begin{aligned} \rho(\boldsymbol{p}, \boldsymbol{r}) &:= \frac{1}{Z_q(\boldsymbol{r})} D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{r}) = -\sum_{i=1}^{n+1} P_i(\boldsymbol{r}) \left(\ln_q(p_i) - \ln_q(r_i)\right) \\ &= \psi(\boldsymbol{\theta}(\boldsymbol{p})) + \psi^*(\boldsymbol{\eta}(\boldsymbol{r})) - \sum_{i=1}^{n} \theta^i(\boldsymbol{p}) \eta_i(\boldsymbol{r}), \end{aligned} \tag{3}$$

where

$$\begin{aligned} \theta^i(\boldsymbol{p}) &:= \ln_q(p_i) - \ln_q(p_{n+1}), \quad \eta_i(\boldsymbol{p}) := P_i(\boldsymbol{p}), \quad i = 1, \cdots, n, \\ \psi(\boldsymbol{\theta}(\boldsymbol{p})) &:= -\ln_q(p_{n+1}), \quad \psi^*(\boldsymbol{\eta}(\boldsymbol{p})) := \frac{1}{\kappa}\left(\frac{1}{Z_q(\boldsymbol{p})} - q\right), \end{aligned}$$

with $\kappa := (1-\alpha^2)/4 = q(1-q)$, $\ln_q(s) := (s^{1-q} - 1)/(1-q)$ for $s \geq 0$ and

$$\eta_i(\boldsymbol{p}) = \frac{\partial \psi}{\partial \theta^i}(\boldsymbol{p}), \quad \theta^i(\boldsymbol{p}) = \frac{\partial \psi^*}{\partial \eta_i}(\boldsymbol{p}), \quad i = 1, \cdots, n. \tag{4}$$

Alpha-Voronoi partitions and pattern classification problems

For given m points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ we define *α-Voronoi regions* on $\mathcal{S}^n$ using the α-divergence as follows:

$$\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | D^{(\alpha)}(\boldsymbol{p}_k, \boldsymbol{p}) < D^{(\alpha)}(\boldsymbol{p}_l, \boldsymbol{p})\}, \quad k = 1, \cdots, m.$$

An *α-Voronoi partition (diagram)* on $\mathcal{S}^n$ is a collection of the α-Voronoi regions and their boundaries.

By the conformal relation between $D^{(\alpha)}$ and ρ in (3) we immediately see that the Voronoi region defined by

$$\mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k) := \bigcap_{l \neq k} \{\boldsymbol{p} \in \mathcal{S}^n | \rho(\boldsymbol{p}_k, \boldsymbol{p}) < \rho(\boldsymbol{p}_l, \boldsymbol{p})\}$$

coincides with $\mathrm{Vor}^{(\alpha)}(\boldsymbol{p}_k)$. Furthermore, we have proved [7] that if we regard the escort probabilities $(P_i(\boldsymbol{p}))$ as a new coordinate system for $\boldsymbol{p}$, we can efficiently compute $\mathrm{Vor}^{(\mathrm{conf})}(\boldsymbol{p}_k)$ by the standard algorithm [10] using the polyhedron envelop for the convex potential function ψ. Hence, the boundary for each α-Voronoi region consists of straight line segments (See the figure 1 and 2).

When we apply the Voronoi partitioning technique to pattern classification problems, flexible choice of the representing point for each Voronoi region is significant. For this purpose we define a *weighted α-centroid* and a formula to calculate it. Given m points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ on $\mathcal{S}^n$ and weights $w_k \geq 0, k = 1, \cdots, m$, the weighted α-centroid $\boldsymbol{c}_w^{(\alpha)}$ is defined by the minimizer of the following problem:

$$\min_{\boldsymbol{p} \in \mathcal{S}^n} \sum_{k=1}^{m} w_k D^{(\alpha)}(\boldsymbol{p}, \boldsymbol{p}_k).$$

By differentiating the above with θ^i and considering the optimality condition, we have the escort probabilities of $\boldsymbol{c}_w^{(\alpha)}$ as

$$P_i(\boldsymbol{c}_w^{(\alpha)}) = \frac{1}{\sum_{k=1}^m w_k Z_q(\boldsymbol{p}_k)} \sum_{k=1}^m w_k Z_q(\boldsymbol{p}_k) P_i(\boldsymbol{p}_k), \quad i = 1, \cdots, n+1.$$

Thus, the weighted α-centroid is represented as the usual weighted average in the escort probabilities.

While the α-Voronoi partition has a one-dimensional freedom in adjusting the parameter α (or equivalently q), the adjustment tends not to work well for large α (or $q \approx +\infty$) for distributions that have the largest probabilities for the same event. This can be understood by considering the corresponding escort distributions, i.e., they are concentrated in the same corner of the simplex in such a situation. Similarly, the classification tends not to work when $\alpha \approx -1$ (or $q \approx 0$) because all the corresponding escort distributions are concentrated near the uniform distribution.

To resolve this problem, we introduce the following projective transformation:

$$\Pi_{\boldsymbol{t}} : \mathcal{S}^n \ni (p_i) \mapsto (\tilde{p}_i) \in \mathcal{S}^n, \quad \text{where } \tilde{p}_i(\boldsymbol{p}) := \frac{t_i^{-1} p_i}{\sum_{j=1}^{n+1} t_j^{-1} p_j}$$

for $\boldsymbol{t} = (t_1, \cdots, t_{n+1})$ in the positive orthant.

Note that the inverse projective transformation is $\Pi_{\boldsymbol{t}^{-1}}$ where $\boldsymbol{t}^{-1} := (t_1^{-1}, \cdots, t_{n+1}^{-1})$, and $\Pi_{\boldsymbol{t}}$ for $\boldsymbol{t} \in \mathcal{S}^n$ maps $\boldsymbol{t}$ to the uniform distribution. Hence, if we use $\Pi_{\boldsymbol{t}}$ as a preconditioner for given data of discrete distributions before the α-Vononoi partitioning, we can expect the inprovement of classification performance.

Illustrative numerical examples

As an illustrative example we consider α-Voronoi partition of the the given four kinds of discrete distributions (the left one of Figure 1) into four classes. In spite of adjusting α (equivalently q) one of data denoted by the symbol $\bullet$ is classified into the region of data denoted by $\triangle$. The right one of Figure 1 is representation of the α-Voronoi partition in escort probabilities.

Executing projective transformation before the partitioning given data are classified successfully (Figure 2). In both figures the symbol $*$ denotes unweighted ($w_1 = \cdots = w_4$) α-centroids for the corresponding classes, and $\boldsymbol{t}$ for the projective transformation is the unweighted α-centroid of those of four data classes.

Concluding remarks

As is summarized in the preliminary section, introduction of the escort probabilities and the conformal divergence enables us to construct α-Voronoi partitions via the efficient standard algorithm involving a convex potential function. We apply this algorithm to pattern classification problems for data expressed by discrete probability distributions.

While the Voronoi partition technique might not be more efficient than the other naive methods in higher dimensional case, the underlying geometrical idea could be applied to other application such as likelihood ratio test.

Bibliography

[1] T. M. Cover and J. A. Thomas, Elements of information theory, Wiley (1991).

[2] H. Imai and M. Inaba, Divergence-Based Geometric Clustering and Its Underlying Discrete Proximity Structures *IEICE Trans. Inf. & Syst.*, **E83-D** (2000) 27.

[3] G. Lebanon, Metric Learning for Text Documents *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **28** (2005) 497.

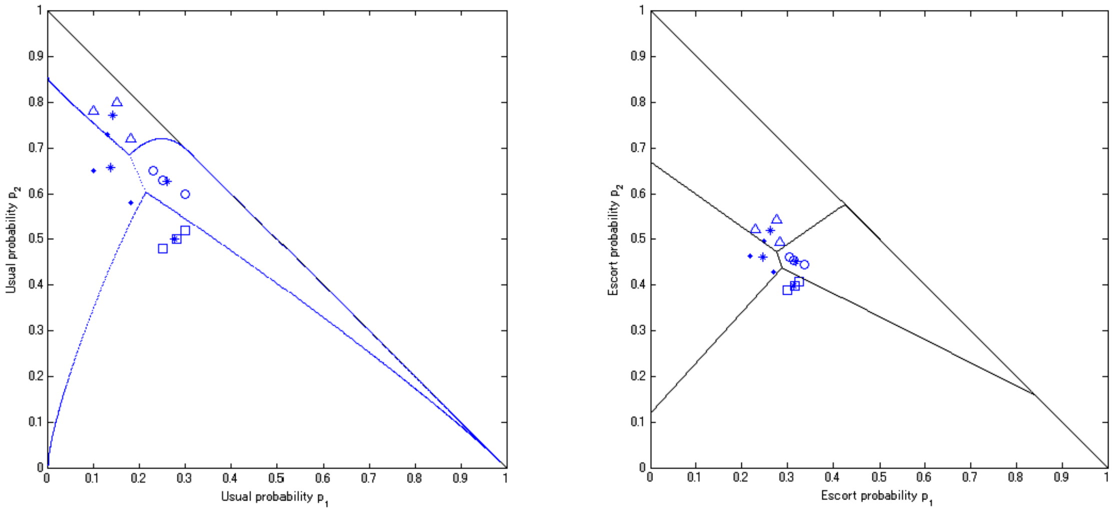

Figure 1: α-Voronoi diagram on $\mathcal{S}^2$ w.r.t. usual probabilities (left) and escort probabilities (right) for q=0.4

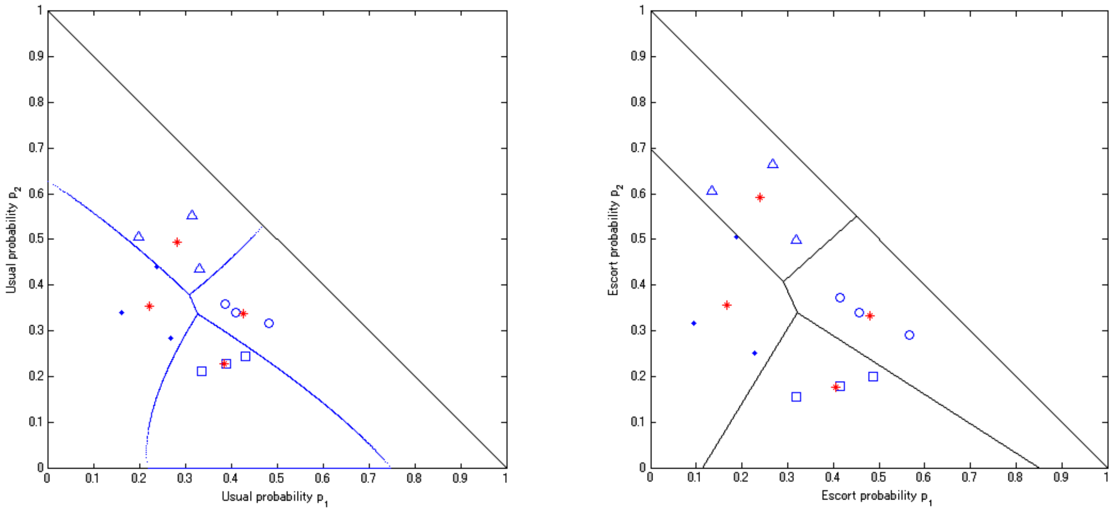

Figure 2: α-Voronoi diagram on $\mathcal{S}^2$ w.r.t. usual probabilities (left) and escort probabilities (right) for q=0.4 with projective transformation as preconditioning

[4] J.-D. Boissonnat, F. Nielsen and N. Nock, *Discrete Comput. Geom.*, **44** (2010) 281.

[5] S.-I. Amari, *Differential-Geometrical Methods in Statistics*, (Springer-Verlag, Berlin, 1985).

[6] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, (AMS&Oxford, Rhode Island, 2000).

[7] A. Ohara, H. Matsuzoe and S.-I. Amari, Conformal geometry of escort probability and its applications, Modern Physics Letters B, Vol. 26, No. 10, 1250063, 2012

[8] A. Ohara, H. Matsuzoe and S.-I. Amari, A dually flat structure on the space of escort distributions, arXiv:1010.4965v1, 2010.

[9] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems*, (Cambridge University Press, Cambridge, 1993).

[10] H. Edelsbrunner, *Algorithms in Combinatorial Geometry* (Springer-Verlag, 1987).

Authors' Information

Atsumi Ohara *- Department of Electrical and Electronics Engineering, University of Fukui, Bunkyo 3-9-1, Fukui 910-8507, Japan; e-mail: ohara@fuee.u-fukui.ac.jp*
Major Fields of Scientific Research: Information Geometry

Yuya Nagatani *- Department of Electrical and Electronics Engineering, University of Fukui, Bunkyo 3-9-1, Fukui 910-8507, Japan; e-mail: nagatani.hase@gmail.com*
Major Fields of Scientific Research: Pattern recognition