Applications of Information Geometry to Audio Signal Processing

Arnaud Dessein* and Arshia Cont

IRCAM, CNRS UMR 9912
1 place Stravinsky, 75004 Paris, France
{arnaud.dessein,arshia.cont}@ircam.fr

In this talk, we present some applications of information geometry to audio signal processing. We seek a comprehensive framework that allows to quantify, process and represent the information contained in audio signals. In digital audio, a sound signal is generally encoded as a waveform, and a common problematic is to extract relevant information about the signal by computing sound features from this waveform. A key issue in this context is then to bridge the gap between the raw signal or low-level features (e.g. attack time, frequency content), and the symbolic properties or high-level features (e.g. speaker, instrument, music genre).

We address this issue by employing the theoretical framework of information geometry. In general terms, information geometry is a field of mathematics that studies the notions of probability and of information by the way of differential geometry [1]. The main idea is to analyze the geometrical structure of differential manifold owned by certain families of probability distributions which form a statistical manifold. We aim to investigate the intrinsic geometry of families of probability distributions that represent audio signals, and to manipulate informative entities of sounds within this geometry.

We focus on the statistical manifolds related to exponential families. Exponential families are parametric families of probability distributions that encompass most of the distributions commonly used in statistical learning. Moreover, exponential families equipped with the dual exponential and mixture affine connections possess two dual affine coordinate systems, respectively the natural and the expectation parameters. The underlying dually flat geometry exhibits a strong Hessian dualistic structure, induced by a twice differentiable convex function, called potential, together with its Legendre-Fenchel conjugate. This geometry generalizes the standard self-dual Euclidean geometry, with two dual Bregman divergences instead of the self-dual Euclidean distance, as well as dual geodesics, a generalized Pythagorean theorem and dual projections.

However, the Bregman divergences are generalized distances that are not symmetric and do not verify the triangular inequality in general. From a computational viewpoint, several machine learning algorithms that rely on strong metric properties possessed by the Euclidean distance are therefore not suitable anymore. Yet, recent works have proposed to generalize some of these algorithms to the case of exponential families and of their associated Bregman

---

* Part of this work was completed while the author was visiting the Japanese-French Laboratory for Informatics, Tokyo, Japan.

divergences [2–6]. It is thus possible, with a single generic implementation, to consider numerous and widely used statistical models or divergences, in algorithms such as centroid computation and hard clustering (k-means), parameter estimation and soft clustering (expectation-maximization), proximity queries in ball trees (nearest-neighbors search, range search).

We discuss the use of this powerful computational framework for applications in audio. The general paradigm is the following. The audio signal is first represented with sound features. We then model these features with probability distributions and apply the tools of information geometry onto these distributions. In particular, it allows to redefine the notion of similarity between two signals in an information setting by employing the canonical divergence of the underlying statistical manifold. This paradigm has been recently investigated for audio data mining in [7]. We show in particular how to segment audio streams into quasi-stationary chunks that form consistent informative entities. These entities can then be treated as symbols for applications such as music similarity analysis, musical structure discovery, query by similarity, audio recombination by concatenative synthesis, and computer-assisted improvisation.

Acknowledgments. This work was supported by a doctoral fellowship from the UPMC (EDITE) and by a grant from the JST-CNRS ICT (Improving the VR Experience).

References

1. Amari, S.-i. and Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Translations of Mathematical Monographs, Vol. 191 (2000)
2. Banerjee, A., Merugu, S., Dhillon, I. S., Ghosh, J.: Clustering with Bregman Divergences. Journal of Machine Learning Research. 6, 1705–1749 (2005)
3. Cayton, L.: Fast Nearest Neighbor Retrieval for Bregman Divergences. In: 25th International Conference on Machine Learning. Helsinki, Finland (2008)
4. Cayton, L.: Efficient Bregman Range Search. Advances in Neural Information Processing Systems. Curran Associates, Inc. 22, 243–251 (2009)
5. Garcia, V., Nielsen, F., Nock, R.: Levels of Details for Gaussian Mixture Models. In: 9th Asian Conference on Computer Vision, pp. 514–525. Xi'an, China (2009)
6. Nielsen, F., Nock, R.: Sided and Symmetrized Bregman Centroids. IEEE Transactions on Information Theory. 55(6), 2882–2904 (2009)
7. Cont A., Dubnov S., and Assayag G.: On the Information Geometry of Audio Streams with Applications to Similarity Computing. IEEE Transactions on Audio, Speech and Language Processing. 19(1), to appear (2011)