RESEARCH PAPER

Anomaly detection in the probability simplex under different geometries

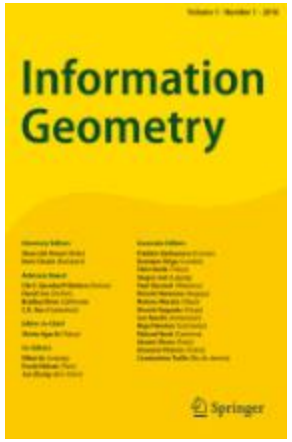
Uriel Legaria[1] · Sergio Mota[1] · Sergio Martinez[1] · Alfredo Cobá[2] · Argenis Chable[2] · Antonio Neme[3]

Received: 4 August 2022 / Revised: 24 April 2023 / Accepted: 13 May 2023 /
Published online: 30 May 2023

Abstract

An open problem in data science is that of anomaly detection. Anomalies are instances that do not maintain a certain property that is present in the remaining observations in a dataset. Several anomaly detection algorithms exist, since the process itself is ill-posed mainly because the criteria that separates common or expected vectors from anomalies are not unique. In the most extreme case, data is not labelled and the algorithm has to identify the vectors that are anomalous, or assign a degree of anomaly to each vector. The majority of anomaly detection algorithms do not make any assumptions about the properties of the feature space in which observations are embedded, which may affect the results when those spaces present certain properties. For instance, compositional data such as normalized histograms, that can be embedded in a probability simplex, constitute a particularly relevant case. In this contribution, we address the problem of detecting anomalies in the probability simplex, relying on concepts from Information Geometry, mainly by focusing our efforts in the distance functions commonly applied in that context. We report the results of a series of experiments and conclude that when a specific distance-based anomaly detection algorithm relies on Information Geometry-related distance functions instead of the Euclidean distance, the performance is significantly improved.

Keywords Anomaly detection · Probability simplex · Information geometry



Volume 6
Issue 2
Article 71

RESEARCH PAPER

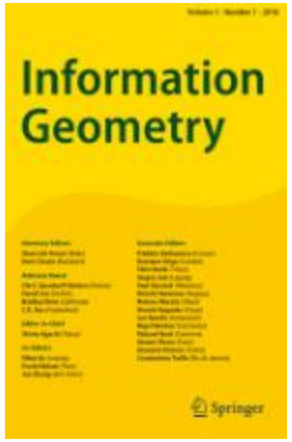
Differential geometry for the optimal design of the contingent valuation method

Hisatoshi Tanaka[1]

Received: 9 August 2022 / Revised: 30 April 2023 / Accepted: 13 May 2023 /
Published online: 31 May 2023

Abstract
The contingent valuation method (CVM) is a widely used experimental method to measure the monetary value of goods. However, CVM estimates are sensitive to experiment design. In this study, we formulated the optimal design problem as a minimization problem of the Fisher information metric of a gradient vector field generated by using the statistical model of the CVM. Furthermore, a necessary and sufficient condition of the optimal design was proven.

Keywords Optimal design · Contingent valuation method · Binary response model · Efficiency bound



Volume 6
Issue 2
Article 72

RESEARCH PAPER

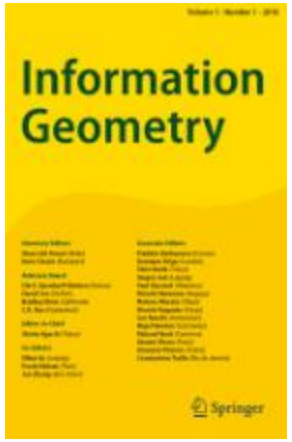
Object embedding using an information geometrical perspective

Taiki Sugiura[1] · Noboru Murata[1]

Received: 2 May 2022 / Revised: 9 June 2023 / Accepted: 20 June 2023 /
Published online: 27 July 2023

Abstract
Acquiring vector representations of objects is essential for applying machine learning, statistical inference, and visualization. Although various vector acquisition methods have been proposed considering the relationship between objects in target data, most of them are supposed to use only a specific relevance level. In real-world data, however, there are cases where multiple relationships are contained between objects, such as time-varying similarity in time-series data or various weighted edges on graph-structured data. In this paper, a vector acquisition method which assigns vectors in a single coordinate system to objects preserving the information given by multiple relations between objects is proposed. In the proposed method, a logarithmic bilinear model parameterized by representation vectors is utilized for approximating relations between objects based on a stochastic embedding idea. The inference algorithm proposed in this study is interpreted in terms of information geometry: the m-projection from the probability distribution constructed from observed relations on the model manifold and the e-mixture in the model manifold are alternately repeated to estimate the parameters. Finally, the performance of the proposed method is evaluated using artificial data, and a case study is conducted using real data.

Keywords Constellation · Relevance · Mapping · Stochastic embedding · Mixture distribution



Volume 6
Issue 2
Article 73

RESEARCH PAPER

Conelike radiant structures

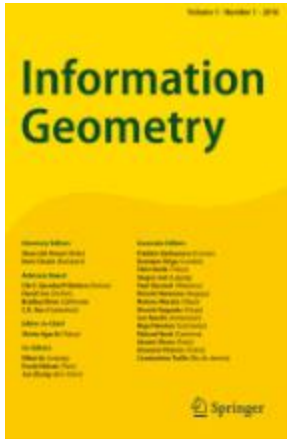
Daniel J. F. Fox[1]

Received: 11 December 2022 / Revised: 9 June 2023 / Accepted: 29 June 2023 /
Published online: 4 October 2023

Abstract

Analogues of the classical affine-projective correspondence are developed in the context of statistical manifolds compatible with a radiant vector field. These utilize a formulation of Einstein equations for special statistical structures that generalizes the usual Einstein equations for pseudo-Riemannian metrics and is of independent interest. A conelike radiant structure is a not necessarily flat affine connection equipped with a family of surfaces that behave like the intersections of the planes through the origin with a convex cone in a real vector space. A radiant structure is a torsion-free affine connection and a vector field whose covariant derivative is the identity endomorphism. A radiant structure is conelike if for every point and every two-dimensional subspace containing the radiant vector field there is a totally geodesic surface passing through the point and tangent to the subspace. Such structures exist on the total space of any principal bundle with one-dimensional fiber and on any Lie group with a quadratic structure on its Lie algebra. The affine connection of a conelike radiant structure can be normalized in a canonical way to have antisymmetric Ricci tensor. Applied to a conelike radiant structure on the total space of a principal bundle with one-dimensional fiber this yields a generalization of the classical Thomas connection of a projective structure. The compatibility of radiant and conelike structures with metrics is investigated and yields a construction of connections for which the symmetrized Ricci curvature is a constant multiple of a compatible metric that generalizes well-known constructions of Riemannian and Lorentzian Einstein–Weyl structures over Kähler–Einstein manifolds having nonzero scalar curvature. A formulation of Einstein equations for special statistical manifolds is given that generalizes the Einstein–Weyl equations and encompasses these more general examples. There are constructed left-invariant conelike radiant structures on a Lie group endowed with a left-invariant nondegenerate bilinear form, and the case of three-dimensional unimodular Lie groups is described in detail.

Keywords Statistical structures · Einstein equations · Radiant manifolds · Projective structures



Volume 6
Issue 2
Article 74