Bernd Hamann
Stratovan

10/16/2020

(NEAR-OPTIMAL) DISTRIBUTION APPROXIMATION

- IDEA: Construct a "high-quality" approximation (and representation) of a probability density/distribution function (PDF) based on the high-dimensional feature points/vectors of the "prototype PDFs" of specific material types/classes of interest — used for comparison with the PDF of a "new material type/class" to be classified.

- CONSIDERATIONS:
  1) Training or prototype data are defined by objects (voxelized) of the same type/class — and creating a discrete, sampled version of the associated high-dimensional PDF of feature values by estimating these feature values for all voxels.
  2) THE SAMPLED PDF — or p-function — IS "SPARSELY SAMPLED" WHEN THE NUMBER OF FEATURES/DIMENSIONS IS "LARGE" — relative to the number of feature points/vectors.

■ DISTRIBUTION APPROXIMATION – Cont'd.

... 3) The goal is to define and effectively compute and represent an ANALYTICAL (smooth, continuous) p-function over the high-dimensional feature space – approximating the unknown PDF of a material type/class in a near-optimal way.

4) High-quality p-function approximations (normalized) can be used to determine a similarity measure for two objects/materials – used for probability-based classification.

5) Each sample (= voxel) of an object of a specific material type/class defines one feature point/vector in high-dimensional space. The density/distribution of all samples representing a specific type/class is used to define and compute an analytical p-function based on a tessellation of the finite set of feature points/vectors.

6) The VORONOI DIAGRAM is used as tessellation. The TILE SIZE of a sample point's associated tile defines the point's p-value.

...

Stratovan

DISTRIBUTION APPROXIMATION – Cont'd.

... 6) ... Specifically, the density value ρ_i associated with sample point $\mathbf{x}_i$ is defined as $\frac{1}{v_i}$.(*) Here, $\mathbf{x}_i$ is a sample point in k-dimensional feature space, i.e., $\mathbf{x}_i = (x_1^i, \ldots, x_k^i)$, $i = 1, \ldots, n$; v_i is the (hyper-) volume (or size) of $\mathbf{x}_i$'s tile.

(*) Two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ can have exactly the same values. Thus: $\rho = \frac{2}{v}$ for this point with multiplicity 2.

7) The desired analytical p-function is obtained via construction of a function $p(\mathbf{x})$ that satisfies the interpolation conditions $p(\mathbf{x}_i) = \rho_i$, $i = 1, \ldots, n$. The SIBSON interpolation method is a "natural method" to interpolate the ρ_i-values for sites $\mathbf{x}_i$ tessellated by a Voronoi diagram.

8) The Voronoi diagram of a set of points is based on DISTANCE and a proper METRIC – to determine, for example, a point's tile and derived quantities. The k dimensions of a feature point/vector can represent quite different properties using various units. THIS ASPECT MUST BE ADDRESSED IN PRACTICE, BUT IT IS NEGLECTED IN THE DISCUSSION.

DISTRIBUTION APPROXIMATION – Cont'd.

... 9) The goal is to compare two p-functions, $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, to determine the degree of similarity of the types/classes of materials. For this purpose p-functions will be normalized, i.e., $\| p(\mathbf{x}) \| = 1$, where

$$\| p(\mathbf{x}) \| = \sqrt{\langle p(\mathbf{x}), p(\mathbf{x}) \rangle} = \sqrt{\int_{\Omega} p(\mathbf{x}) \cdot p(\mathbf{x}) \, d\mathbf{x}} \, .$$

10) Two p-functions $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, both normalized, will be compared via the measure

$$d(p_1(\mathbf{x}), p_2(\mathbf{x})) = \| p_1(\mathbf{x}) - p_2(\mathbf{x}) \| \, .$$

A value of $d = 0$ indicates that two materials are identical — per available data – and a relatively increasing d-value indicates increasing dissimilarity of materials. FOR CLASSIFICATION PURPOSES, IT IS POSSIBLE TO CONCLUDE THAT TWO MATERIALS ARE "THE SAME" WHEN $d < \varepsilon$, OR THAT THEY ARE OF SIMILARITY d.

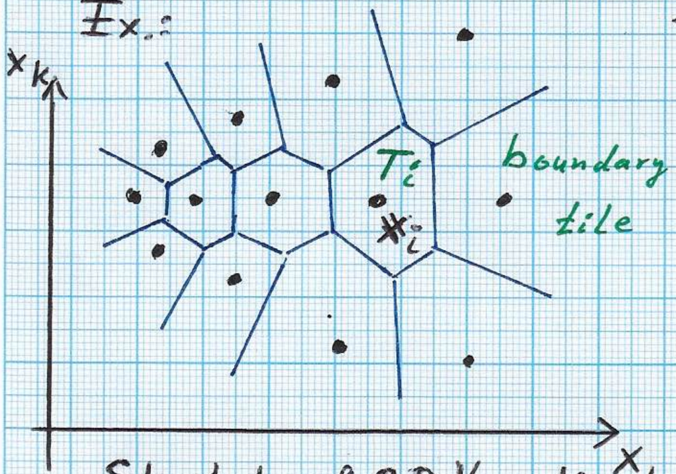
Bernd Hamann

10/24/2020

Stratovan

■ DISTRIBUTION APPROXIMATION – Cont'd.

Ex.:



Sketch of 2D Vor. dia.

→ SITES $x_i = (x_1^i, \dots, x_k^i)$, $i = 1, \dots, n$, shown as '•', define a unique VORONOI DIAGRAM consisting of (parts of) PERPENDICULAR BISECTORS, shown as '——'.

→ The TILE T_i, associated with site x_i, has a (hyper-) volume called v_i. The value of v_i can be finite or infinite (in the case of boundary tiles).

→ Multiple samples $x_i, x_j, x_k, \dots$ can have the same values for all feature dimensions. In this case the one unique site position/location in the Voronoi diagram has a MULTIPLICITY > 1, and the multiplicity value is defined by the number of samples having exactly the same feature values for all dimensions.

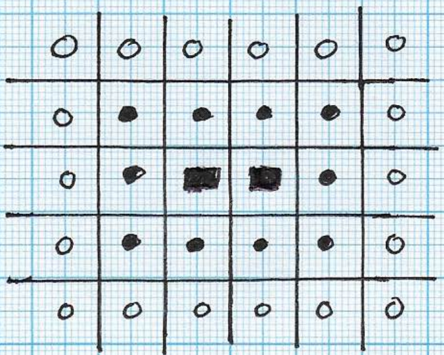
→ The DENSITY VALUE ρ_i of a site is defined as m_i/v_i, assuming the site's tile has FINITE VOLUME v_i; ρ_i is defined as ε when a site's tile has INFINITE VOLUME v_i (boundary tiles).

Bernd Hamann
Stratovan

10/25/2020

DISTRIBUTION APPROXIMATION – Cont'd.

Ex.:



define density values ρ_i

ε	ε	ε	ε	ε	ε
ε	1	1	1	1	ε
ε	1	2	2	1	ε
ε	1	1	1	1	ε
ε	ε	ε	ε	ε	ε

P-function via Sibson interpolation (normalized)

- Given are 30 unique sites (in this case lying on a regular Cartesian mesh).
- Sites shown as 'o' or '•' have multiplicity 1; sites '■' have multiplicity 2.
- The volume (= area) of all tiles with finite volume is 1.
- In this example, the computed density values are:

  2 – interior sites with multiplicity 2
  1 – interior sites with multiplicity 1
  ε – all boundary sites

- Construct the Sibson interpolation function for the domain of interest; define the P-function via normalization of the Sibson interpolation function.

~ BH

• EFFICIENT IMPLEMENTATION NECESSARY!