ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50

Robust Topology Representing Networks

Michaël Aupetit*
CEA - DASE - BP 12 - 91680 - Bruyères-Le-Châtel - France

Abstract.
Martinetz and Schulten proposed the use of a Competitive Hebbian Learning (CHL) rule to build Topology Representing Networks. From a set of units and a data distribution, a link is created between the first and second closest units to each datum, creating a graph which preserves the topology of the data set. However, one has to deal with finite data distributions generally corrupted with noise, for which CHL may be unefficient. We propose a more robust approach to create a topology representing graph, by considering the density of the data distribution.

1 Introduction

In many applications [1, 7], it is intended to represent the topology of a manifold[1] $\mathcal{M} \subseteq E$ only known through a finite set $\underline{v}$ of samples in a bounded domain $E \subset \mathbb{R}^D$. $\mathcal{M}$ is the support of the p.d.f from which the samples are drawn.

A way to do this is to define a set $\underline{w} \in E$ of N units, which tend to represent the data distribution after a prior Vector Quantization (VQ) phase [6]. From $\underline{w}$, a set of links may be built to represent the topology of $\mathcal{M}$, according to the following idea [3, 7]: each unit w_i is representative of the data for which it is the closest unit among all the units of $\underline{w}$. All these data fall in a region $\mathcal{M}_i$ which is the intersection between the manifold $\mathcal{M}$ and the Voronoï region $\mathcal{V}_i$ of $w_i \in \underline{w}$ defined as : $\mathcal{V}_i = \mathcal{V}_{(E,\underline{w})}(w_i) = \{v \in E \mid \forall w_j \in \underline{w}, \|v - w_i\| \leq \|v - w_j\|\}$.

The only pieces of $\mathcal{M}$ which may be possibly connected to $\mathcal{M}_i$ are the $\mathcal{M}_j$ for which $\mathcal{V}_j$ shares a common boundary with $\mathcal{V}_i$, *i.e.* such that w_j is a natural (or Delaunay) neighbor of w_i. If we construct DT($\underline{w}$) the Delaunay triangulation [4] of $\underline{w}$ defined as the set of links l_{ij} which connect natural neighbor units w_i and w_j whose Voronoï regions share a common boundary,

$$\mathrm{DT}(\underline{w}) = \{l_{ij} \subseteq \underline{w} \mid \mathcal{V}_i \cap \mathcal{V}_j \neq \emptyset\} \text{ with } l_{ij} = \{w_i, w_j\} \quad (1)$$

then the set IDT($\underline{w}, \mathcal{M}$) of links which best represent the topology of $\mathcal{M}$, is a subset of DT($\underline{w}$) (IDT stands for "Induced Delaunay Triangulation" [7]) .

*aupetit@dase.bruyeres.cea.fr

[1] In this paper, $\mathcal{M}$ is called a manifold for short, it is in general, a collection of manifolds connected or not, which may have various intrinsic dimensions.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50

The main problem which arises is now how to build $\mathrm{IDT}(\underline{w}, \mathcal{M})$ of topology representing links having in mind that $\mathcal{M}$ is not given but partly through $\underline{v}$.

Edelsbrunner and Shah [3] proposed that a link is drawn between two natural neighbor units for which the intersection between $\mathcal{M}$ and their common Voronoï boundary is not empty: $\mathrm{IDT}_{\mathrm{E\&S}}(\underline{w}, \mathcal{M}) = \{l_{ij} \subseteq \underline{w} \mid \mathcal{V}_i \cap \mathcal{V}_j \cap \mathcal{M} \neq \emptyset\}$. However, the analytical expression of $\mathcal{M}$ must be known to test this condition[2].

Martinetz and Schulten [7] present the Topology Representing Network (TRN) and later under different assumptions Bruske and Sommer [2], the Optimally Topology Preserving Map (OTPM), which both use the Competitive Hebbian Learning (CHL) to build $\mathrm{IDT}_{\mathrm{CHL}}(\underline{w}, \underline{v})$. They consider each datum of $\underline{v}$, searching its closest and second closest units in $\underline{w}$, and then creating a link between both units: $\mathrm{IDT}_{\mathrm{CHL}}(\underline{w}, \underline{v}) = \{l_{ij} \subseteq \underline{w} \mid \exists v \in \underline{v}, v \in \mathcal{V}_{ij}\}$, where $\mathcal{V}_{ij}$ is called the 2^{nd}-order Voronoï region associated to $\{w_i, w_j\}$ with $\mathcal{V}_{ij} = \left\{v \in E \mid v \in \mathcal{V}_{(E,\underline{w}\setminus w_i)}(w_j) \cap \mathcal{V}_{(E,\underline{w}\setminus w_j)}(w_i)\right\}$ (Figure 1a).

CHL is the only practical algorithm known to build an IDT from a finite set of samples. However, it is prone to a series of limits:

- The connectivity of the manifold between two units is inferred from the occurence of a single datum. It should be more robust to infer it from the density of the data in the region of influence of the corresponding link, which is here the 2^{nd}-order Voronoï region.
- If the data set is corrupted with noise, the topology of the support manifold including the noise will be represented, while it would be more interesting to represent the topology of the principal manifolds, *i.e.* the support manifold of the data distribution if there were no noise[3].
- The 2^{nd}-order Voronoï regions may have no intersection at all with their corresponding edge. In other words, a dense connected manifold containing both extreme units of that edge (*e.g.* a straight line between these units) cannot give rise to it: there is no self-consistency property.
- The desired IDT, as subgraph of DT, may define a set of k-simplices ($k \in 0 \ldots N-1$) which then represent k-dimensional pieces of $\mathcal{M}$. However, CHL cannot generates 0-manifolds[4] represented by lonely units, because for $N \geq 2$, there always exists a first and a second closest units to a datum, which therefore are linked together.

We propose to solve these problems by considering a two-phase approach:

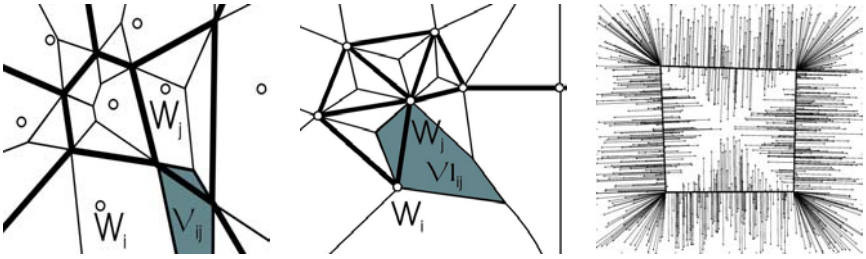
- First, a Topological Graph $\mathrm{TG}(\underline{w}, \underline{v}) \subseteq \mathrm{DT}(\underline{w})$ is created.
- Second, a pruning process of TG takes place according to a statistical criterion depending on the empirical p.d.f of the data distribution. The resulting graph is called $\mathrm{IDT}_{\mathrm{robust}}(\underline{w}, \underline{v})$.

[2]The Lebesgue measure of the intersection $\mathcal{V}_i \cap \mathcal{V}_j$ is 0, hence no finite sampling can be useful to test the condition in practice.

[3]Assuming this noise is zero-mean Gaussian additive

[4]k-manifold stands for k-dimensional manifold

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50



(a) (b) (c)

Figure 1: (a) Boundaries of Voronoï (bold lines) and 2nd-order Voronoï (thin lines) regions of a set of units (circles). (b) Gabriel Graph (bold lines) and boundaries of Voronoï regions (thin lines) of its edges and vertices. (c) Example of data projection.

2 Building a topological graph

Computing DT($\underline{w}$) is known to be an intractable task for even quite small D-dimensional space (computation in time $O(N^{\lceil \frac{D}{2} \rceil})$ [4])[5].

In general $D \gg 3$, so we need a way to approximate DT($\underline{w}$) to get the maximum number of links likely to be relevant for representing the topology of $\mathcal{M}$, before pruning irrelevant ones.

The IDT_{CHL} is an appealing solution which generally has more topological links than needed especially when data set is noisy. But it may miss some links due to the too small size of the 2nd-order Voronoï regions in near degenerate cases (Figure 2c-d). The Gabriel Graph (GG) [5] is a subgraph of DT($\underline{w}$) which may be built in time $O(D.N^3)$: $\text{GG}(\underline{w}) = \{l_{ij} \subseteq \underline{w} | \forall\ w_k \in \underline{w}, d_{ki} + d_{kj} > d_{ij}\}$ with $d_{ab} = (w_a - w_b)^2$. $GG(\underline{w})$ allows to get some of the links missed by IDT_{CHL}, because it does not depend on the density of $\underline{v}$, but it may miss some links too because it cannot create any obtuse or right triangles of DT($\underline{w}$) (Figure 2a-b).

We propose to define: $\text{TG}(\underline{w}, \underline{v}) = \text{GG}(\underline{w}) \cup \text{IDT}_{\text{CHL}}(\underline{w}, \underline{v})$.

3 Pruning the topological graph

3.1 Region of influence and projection

In order to obtain the self-consistency property, we propose to set the region of influence of the links as their Voronoï region $\mathcal{V}_{l_{ij}}$ instead of $\mathcal{V}_{ij}$(Figure 1b): $\mathcal{V}_{l_{ij}} = \mathcal{V}_{E,\underline{L},\underline{w}}(l_{ij}) = \{v \in E \mid \forall x \in \{\underline{L}, \underline{w}\}, d(v,x) \geq d(v, l_{ij})\}$ with

$$d(v,x) = \begin{cases} (v-x)^2\ , \text{ if } x \in \underline{w} \\ (v-v_p)^2, \text{ if } x = l_{ab} \in \underline{L} \text{ with } v_p = w_a + k_{ab}(v) \cdot (w_b - w_a) \\ \quad \text{where } k_{ab}(v) = \dfrac{\langle v - w_a | w_b - w_a \rangle}{(w_b - w_a)^2} \text{ and } k_{ab} \in [0,1] \end{cases}$$

[5] CHL does not need the prior construction of DT($\underline{w}$) avoiding this difficulty.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50

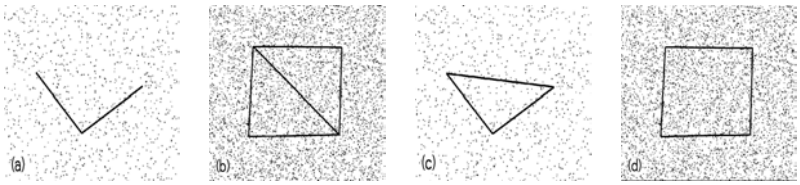


Figure 2: GG misses (a) (finds (b)) a link that IDT_{CHL} finds (c) (misses (d)).

where $\underline{L}$ is a set of links l_{ij} between w_i and w_j, and $\langle .|. \rangle$ denotes the inner product. Thus, each link is entirely inside its Voronoï region. A link represents the piece of $\mathcal{M}$ closest to it than to any other link or unit: it is the bounded linear 1-manifold which best represents this part of $\mathcal{M}$ in the sense of the $\mathcal{L}^2$ distortion measure[6].

3.2 A statistical criterion

Now, we consider all the data which project onto a link (Figure 1c) and we propose a simple hypothesis to test if this link is worth being kept in $\text{TG}(\underline{w}, \underline{v})$.

A link which does not receive any data in its Voronoï region must be cleared, exactly as the "dead" units in the VQ framework. Moreover, it is possible to refine the selection criterion by focusing on the distribution of the projections onto the link: if there are more data which project on the middle part of the link than near its both extremes, it is supposed that the underlying support manifold has no "hole" in that place, hence is connected.

Therefore, we compute a 3-bin histogram of the projections :

$$\forall l_{ij} \in \text{TG}(\underline{w}, \underline{v}), \begin{cases} h_1(l_{ij}, \sigma) = \text{card}(\{v \in \{\underline{v} \cap \mathcal{V}_{l_{ij}}\} \mid k_{ij}(v) \leq t_{12}(\sigma)\}) \\ h_2(l_{ij}, \sigma) = \text{card}(\{v \in \{\underline{v} \cap \mathcal{V}_{l_{ij}}\} \mid t_{12}(\sigma) < k_{ij}(v) \leq t_{23}(\sigma)\}) \\ h_3(l_{ij}, \sigma) = \text{card}(\{v \in \{\underline{v} \cap \mathcal{V}_{l_{ij}}\} \mid t_{23}(\sigma) < k_{ij}(v)\}) \end{cases} \tag{2}$$

with $t_{12}(\sigma) = \frac{1}{2}(1-\sigma)$ and $t_{23}(\sigma) = \frac{1}{2}(1+\sigma)$. The link l_{ij} is kept if the bin h_2 contains more data than the others:

$$\text{IDT}_{\text{robust}}(\underline{w}, \underline{v}, \sigma) = \text{TG}(\underline{w}, \underline{v}) \backslash \{l_{ij} \subseteq \text{TG}(\underline{w}, \underline{v}) | h_2(l_{ij}, \sigma) < \max(h_1(l_{ij}, \sigma), h_3(l_{ij}, \sigma))\} \tag{3}$$

with $\sigma \in [0, 1]$ a global parameter allowing to tune the width of the middle bin h_2 for every links, hence the sensitivity of the pruning condition (3). For $\sigma = 0$, all the links are cleared, while for $\sigma = 1$ they are all kept whatever $\underline{v}$.

3.3 A heuristic to set the parameter σ

We propose a heuristic to set σ, by counting the total number $N_l(\sigma)$ of links in $\text{IDT}_{\text{robust}}(\underline{w}, \underline{v}, \sigma)$ according to σ. Experimentally, a relevant σ is $\sigma^* \in [\frac{1}{3}, \frac{2}{3}]$ closest to $\frac{1}{3}$, for which $N_l(\sigma)$ reaches the largest plateau (Figure 3).

[6] in the same way as each unit $w_i \in \underline{w}$ is the 0-manifold which best represents $\mathcal{M}_i$.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50

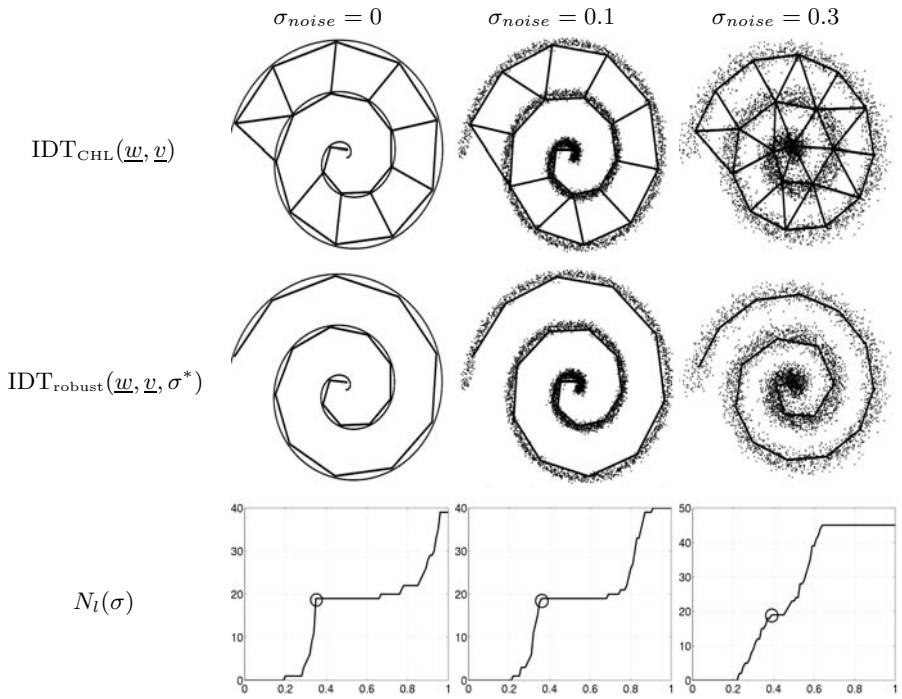


Figure 3: Comparison between $\mathrm{IDT}_{\mathrm{CHL}}$ and $\mathrm{IDT}_{\mathrm{robust}}$ on the spiral data set. The beginning of the largest plateau (circle) in $N_l(\sigma)$ for $\sigma \in [\frac{1}{3}, \frac{2}{3}]$ gives σ^*.

4 Experiments

4.1 Noisy spiral data

We generate a 2-dimensional spiral distribution of 5000 data and add zero-mean Gaussian noise with variance σ^2_{noise}. We place 20 units using a Neural-Gas [6], and compare $\mathrm{IDT}_{\mathrm{CHL}}(\underline{w}, \underline{v})$ *vs* $\mathrm{IDT}_{\mathrm{robust}}(\underline{w}, \underline{v}, \sigma^*)$ (Figure 3).

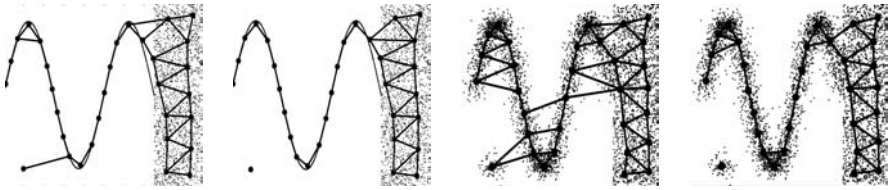
$\mathrm{IDT}_{\mathrm{CHL}}$ fails to represent the topology of the data set even without noise, while $\mathrm{IDT}_{\mathrm{robust}}$ succeed in that task even with large σ_{noise}. The heuristic for choosing σ^* keeps being relevant for σ_{noise} up to 0.3.

4.2 Noisy complex data

We generate a 2-dimensional distribution of 5000 data containing a 0-manifold and a non-linear 1-manifold connected to a 2-manifold, and add zero-mean Gaussian noise with variance σ^2_{noise}. We place 32 units using a Neural-Gas and compare $\mathrm{IDT}_{\mathrm{CHL}}(\underline{w}, \underline{v})$ *vs* $\mathrm{IDT}_{\mathrm{robust}}(\underline{w}, \underline{v}, \sigma^*)$ (Figure 4).

$\mathrm{IDT}_{\mathrm{robust}}$ is again more robust than $\mathrm{IDT}_{\mathrm{CHL}}$ facing pure or noisy data. $\mathrm{IDT}_{\mathrm{robust}}$ is also able to represent 0-manifolds which is impossible for $\mathrm{IDT}_{\mathrm{CHL}}$.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 45-50



(a)IDT_{CHL} $\sigma_{noise}=0$ (b)$\text{IDT}_{\text{robust}}$ $\sigma^*=0.4$ (c)IDT_{CHL} $\sigma_{noise}=0.03$ (d)$\text{IDT}_{\text{robust}}$ $\sigma^*=0.45$

Figure 4: $\text{IDT}_{\text{robust}}$ is less sensitive to noise and can represent 0-manifolds.

5 Conclusion

We present a new way to build Topology Representing Networks by creating a topological graph and pruning this graph by taking into account the density of the data which project onto each edge. This approach is less sensitive to noise than the previous one proposed by Martinetz and Schulten [7], allowing to closer represent the topology of the principal manifolds than that of the support manifolds. It has the self-consistency property, and can represent 0-manifolds.

Experiments should be done in spaces of higher dimension and future researches could be finding an even more efficient statistical criterion than the one based on a 3-bin histogram we test. The definition of topology preservation used in [2, 3, 7] should also be reconsidered in the light of this new approach.

References

[1] Aupetit, M., Couturier, P. & Massotte, P. (2001) Induced Voronoï kernels for principal manifolds approximation. *Advances in self-organizing maps.* N. Allison, H. Yin, L. Allinson, & J. Slack eds, Springer, 54-60.

[2] Bruske, J. & Sommer, G. (1998) Intrinsic Dimensionality Estimation With Optimally Topology Preserving Maps. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **20**(5):572-575.

[3] Edelsbrunner, H. & Shah N.R. (1997) Triangulating Topological Spaces. *International Journal of Computational Geometry and Applications,* **7**(4):365-378.

[4] Fortune, S. (1992) Voronoï diagrams and Delaunay triangulations. *Computing in Euclidean geometry.* D.Z. Du, F. Hwang eds, World Scientific, 193-233.

[5] Gabriel, K.R. & Sokal, R.R.(1969) A new statistical approach to geographic variation analysis. *Syst. Zoology,* **18**:259-278.

[6] Martinetz,T.M., Berkovitch, S.G. & Schulten, K.J. (1993) "Neural-Gas" Network for Vector Quantization and its Application to Time-Series Prediction. *IEEE Trans. on Neural Networks,* **4**(4):558-569.

[7] Martinetz, T.M. & Schulten, K.J. (1994) Topology Representing Networks. *Neural Networks,* **7**(3):507-522.