

Techniques d'apprentissage

Panchalingamoorthy Gajenthiran

Université Paris 8

A Clonal Selection Algorithm with Levenshtein Distance based Image Similarity in Multidimensional Subjective Tourist Information and Discovery of Cryptic Spots by Interactive GHSOM

Introduction

- MPPS
- Classification des données
- GHSOM, Levenshtein et CSAIM

Plan

- Datasets
- Apprentissage
- Expérimentations

Datasets

- Fichier .jpeg
- Nom du lieu
- Évaluation (/4)
- Commentaire

Apprentissage

- Interactive GHSOM
- Distance de Levenshtein
- CSAIM

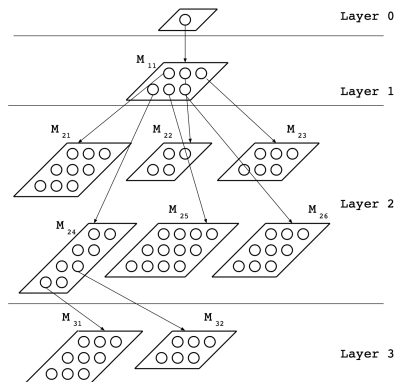
SOM

- Réseau de neurones
- Carte de Kohonen
- Partitionnement de données

SOM

- Taille statique
- Relations hiérarchiques entre les données
- Solution: GHSOM

GHSOM



Architecture du modèle GHSOM

GHSOM

- 1 noeud
- τ_m = architecture du modèle
- τ_u = qualité de la représentation des données
- $mqe_{u_0} = \frac{||w_0 - x||}{n}$

n = nombre de données en entrée

x = donnée en entrée

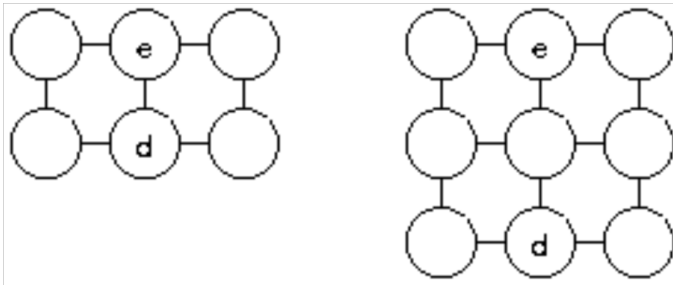
w_0 = vecteur de poids du neurone 0

GHSOM

- BMU
- $w_{i_v}(t+1) = w_{i_v}(t) + \theta(i_{BMU}, i_v, t) * \alpha(t) * (x - w_{i_v})$
 θ = fonction de voisinage
 α = coefficient d'apprentissage
- $mqe_{m_j} = \frac{\sum_{i=0}^n mqe_{u_i}}{u}$
 u = nombre de neurones

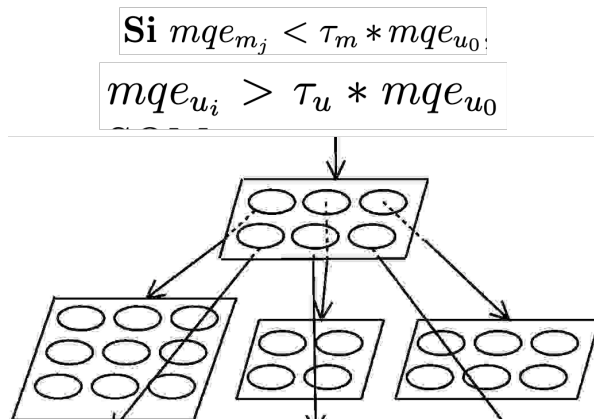
GHSOM

$$\text{Si } mqe_{m_j} \geq \tau_m * mqe_{u_0}$$



Insertion de neurones dans une carte, dans GHSOM

GHSOM



Insertion de neurones dans une carte, dans GHSOM

IGHSOM

- limiter développement des couches
- contrôler τ_m et τ_u

IGH SOM

- $n_{BMU} \leq \delta * n_I$ et $mqe_{m_j} < \tau_m * mqe_{u_0}$
 n_{BMU} = nombre de données en entrée pour le BMU
 n_I = nombre de données en entrée total
 δ = une constante

IGHSOM

- $mqe_{u_i} > \beta * \tau_m * \sum mqe_{u_j}, j \in S_k$ et
 $mqe_{u_i} > \tau_u * mqe_{u_0}$
 $S_k =$ ensemble des BMU

LD

- Distance entre deux chaînes
- Reconnaissance et similarité des images

$$lev_{s,t}(i,j) = \begin{cases} \max(i,j) & , \text{ if } \min(i,j) = 0 \\ \min \begin{cases} lev_{s,t}(i-1,j) + 1, \\ lev_{s,t}(i,j-1) + 1, \\ lev_{s,t}(i-1,j-1) + [s_i \neq t_j] \end{cases} & , \text{ otherwise} \end{cases}$$

Algorithme de la distance de Levenshtein

LD

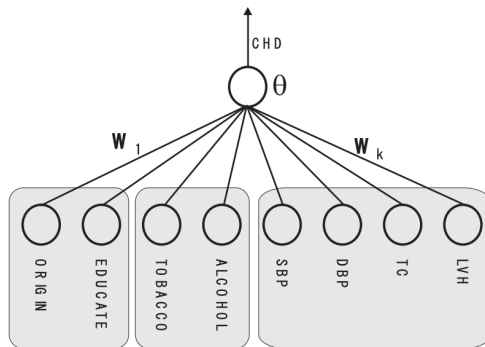
		H	Y	U	N	D	A	I
	0	1	2	3	4	5	6	7
H	1	0	1	2	3	4	5	6
O	2	1	1	2	3	4	5	6
N	3	2	2	2	2	3	4	5
D	4	3	3	3	3	2	3	4
A	5	4	4	4	4	3	2	3

Exemple de l'utilisation de *LD* entre deux chaînes *HYUNDAI* et *HONDA*

CSAIM

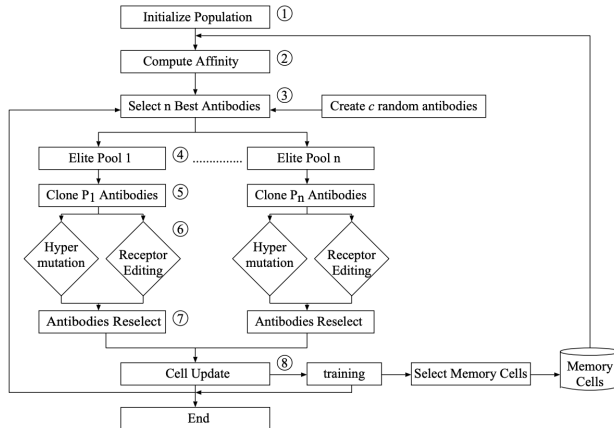
- Interactions entre Abs et Ags
- Selection clonale, HM et RE
- Classification des images

CSAIM



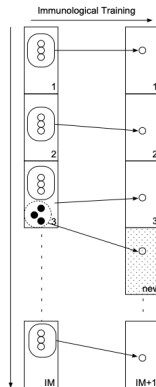
Structure d'un anticorps

CSAIM



Architecture du modèle CSAIM

CSAIM



Clustering des cellules mémoires

Expérimentations

- TF-IDF
- Deux groupes: avec et sans photo
- Pertinence des recherches (DuckduckGo)



Sans photo





Classification de Miyajima sans photo



Avec photos

Including Famous Symbol, Torii




Not Including Famous Symbols

Including the famous keyword of sightseeing spots and goods

High TF-IDF

No Keywords

Low TF-IDF

Classification de Miyajima avec photos

Conclusion

- Modèle robuste/efficace
- Polyvalence
- Manque de données d'apprentissage

Sources

- Shin Kamada Takumi Ichimura. A Clonal Selection Algorithm with Levenshtein Distance based Image Similarity in Multidimensional Subjective Tourist Information and Discovery of Cryptic Spots by Interactive GHSOM. Thèse en informatique, Faculty of Management and Information Systems, Prefectural University of Hiroshima, 2018.
- Andreas Rauber. The ghsom architecture and training process. Technical report.

Sources

- Shin Kamada Takumi Ichimura. Clustering and Retrieval Method of Immunological Memory Cell in Clonal Selection Algorithm. Thèse en informatique, Faculty of Management and Information Systems, Prefectural University of Hiroshima, 2018.
- Merriam Park Software Michael Gilleland. Levenshtein distance. Technical report.