

**Madalina Olteanu**

Professeure des universités en science des données

CEREMADE UMR 7534

Université Paris Dauphine PSL

Pl. du Maréchal de Lattre de Tassigny, 75016 Paris

[olteanu@ceremade.dauphine.fr](mailto:olteanu@ceremade.dauphine.fr)

Gif sur Yvette, le 2 juin 2023

## Remarques ou questions sur le rapport de thèse

### Chapitre 1

- Plusieurs méthodes décrites dans ce chapitre sont basées sur la transmission des positions dans la carte ; il est dit d'ailleurs que « la position des BMU exploite totalement l'aspect topologique d'une SOM ». A mon avis, ceci n'est vrai que si le dépliement de la carte est correct. Dans quelle mesure il peut y avoir des problèmes si la carte n'est pas bien dépliée (données très bruitées, dimension latente trop grande, ...) ? Une discussion, même courte, serait peut être intéressante.
- Les architectures décrites ici ne contiennent que des cartes SOM. Y a-t-il des approches hybrides qui combinent des cartes avec d'autres méthodes (réduction de dimension par auto-encodeurs, par exemple) qui seraient des alternatives à des méthodes comme DSOM, par exemple ? Même si le but n'est pas de les décrire, dire si elles existent ouvrirait plus largement l'état de l'art.
- Page 3 - préciser qu'il s'agit de la version online de l'algorithme, qui existe aussi bien en batch, avec une logique d'apprentissage différente.
- Figure 1.3, page 4 - les données ont implicitement une distribution uniforme, sinon l'aspect de la carte serait différent.
- La phrase « la création d'architectures pour le traitement des données multi-modales se rapproche de la question du traitement des séquences » n'est vraie que partiellement, en raison des hypothèses sur les structures de dépendance qui peuvent être très différentes (causales ou non-causales pour les données séquentielles)
- Equation en haut de la page 34 - il manque une norme dans le dernier terme ou alors les données sont supposées uni-variées ?

### Chapitre 2

- Page 42,  $\omega_e \in \mathcal{D}$  et non dans  $D$ .
- Les références des équations etc ne sont pas toujours les bonnes. Par exemple, page 49, la référence à l'équation 2.1 pointe vers le chapitre 1.
- Etape 2b, page 51, la formule de la mise à jour n'est pas tout à fait correcte ? Il faut reprendre celle de l'algorithme 1, page 55, en la corrigeant d'ailleurs aussi pour qu'elle tienne compte des composantes du vecteur  $\Pi$  (si carte en dimension 2 ou plus).
- Algorithme 1 - à mon avis, il serait plus lisible peut être de préciser en entrée les différents hyper-paramètres.
- Dans la table 2.1, page 56, il manque le  $\sigma^2$ , le nombre d'itérations de l'algorithme global, et  $K$ .

## Chapitre 4

- Dans la Figure 4.1, il n'y a pas du tout du bruit. Est-ce réaliste et/ou suffisamment compliqué comme exemple ?
- Pour quoi avoir modélisé les données, et surtout les BMU comme des variables aléatoires ? Qu'est-ce qu'apporte de plus cette hypothèse par rapport à une relation fonctionnelle ? D'autant plus que dans la plupart des exemples il n'y a pas de bruit. Quel type d'échantillonnage est réalisé ici, si variables aléatoires ? Il me semble qu'il y a des petits points à clarifier ici.
- La carte a 500 noeuds, mais il n'est pas dit quelle est la taille des données d'apprentissage ?
- Section 4.1.1 - la comparaison avec des cartes 1D non-connectée me paraît moyennement judicieuse. Il est clair que cela ne marchera pas, car on casse la structure de dépendance. Il me semble qu'une carte 1D avec une entrée 2D serait plus indiquée comme concurrente ?
- Section 4.3.3. et figure 4.8. - la représentation implique finalement la connaissance de la fonction qui définit la variété. En pratique ceci est impossible, d'autant plus qu'on dispose d'un échantillon, potentiellement bruité. Comment faire pour le cas des données réelles ?

## Chapitre 5

- Section 5.2.1 - combien de points ont été tirés pour une architecture à 500 noeuds ?
- Configuration C : c'est plutôt  $\cos(2\pi X^1)$ , non ?
- Graphique 5.3 - les ordres de grandeur des  $\omega_c$  et  $\omega_e$  semblent assez différents ; serait-il possible de modifier les graphiques pour avoir les  $\omega_c$  sur la même image, et même chose pour les  $\omega_e$  ? Quand est-ce qu'on décide, en pratique, que l'algorithme a convergé ?
- Figure 5.7 - il semblerait qu'il y ait des effets de bord sur l'entrée contextuelle. Une explication ?
- Dans les architectures à trois cartes, il est dit « chaque carte encode la totalité du modèle  $U$  et non seulement son entrée externe ». De mon point de vue, le modèle est la relation fonctionnelle qui relie  $U$  aux entrées,  $U$  en soi n'a aucun intérêt ?
- Sur la prédiction de données manquantes, à quel point est-ce plus puissant que d'autres modèles d'imputation ( $k$ -nn, etc ...). Dans quelle mesure peut-on faire de l'imputation multiple pour gérer l'incertitude ?

## Chapitre 6

- Page 133 - il serait peut être intéressant de mentionner ici que l'information mutuelle n'est autre qu'une divergence KL entre la loi jointe et le produit des lois marginales, d'où le fait qu'elle vaut 0 si indépendance.
- Figure 6.9 - le résultat est étonnant, les cartes en 1 dimension sans information additionnelle font mieux. Cela pose en effet problème pour l'utilisation de l'indicateur. Ce serait à mon avis intéressant d'en discuter plus ce point.

## Chapitre 7

- Combien de points ont été considérés pour les données d'entraînement ?
- Page 158, premier paragraphe - il me semble qu'il y a un problème avec la définition du taux de convergence ?