Robustesse d'une évaluation intégrée

Application des méthodes d'approximation d'intégrales aux évaluation des performances d'Eco-quartiers

May 14, 2014

Motivation

Un même “fait urbain” ([Mangin and Panerai, 1999]) peut avoir plusieurs réalisations à différents niveaux et sous différentes formes.

Q : comment avoir un référentiel global pour toutes les évaluations des réalisations ; et définir une robustesse des évaluations par rapport à celui-ci ? Si possible en amont de la définition des indicateurs (robustesse structurelle) ?

Pistes

- Prise en compte robustesse statistiques par rapport aux données du cas, ainsi que de la fiabilité du nombre de dimensions prises en compte.

- Multi-Criteria Decision Analysis essentielle pour le développement durable [Wang et al., 2009] : pondération des critères pour une approche aggrégée. On se placera dans le cadre d'une aggrégation.

- Cadre : travaux statistiques récents [Varet, 2010] pour l'approximation d'intégrales appliquables à de nombreux domaines, permettant de combiner les deux approches précedentes.

Formulation

$(S_i)_{1\leq i\leq N}$ systèmes territoriaux disjoints (les "éco-quartiers") représentés par des jeux de données et des indicateurs intermédiaires :

$$S_i = (\mathbf{X}_i, \mathbf{Y}_i) \in \mathcal{X}_i \times \mathcal{Y}_i$$

Avec $\mathcal{X}_i = \prod_k \mathcal{X}_{i,k}$ et $\mathcal{X}_{i,k} = \mathbb{R}^{n^X_{i,k} p^X_{i,k}}$ (idem pour les $\mathcal{Y}_i$) et $I_X(i,k)$ (resp. $I_Y(i,k)$) fonction de désignation du type de variable (resp. de l'indicateur)

Definition

L'espace partiel caractéristique du fait urbain est $(\mathcal{X},\mathcal{Y}) \underset{def}{=} (\prod \tilde{\mathcal{X}}_c) \times (\prod \tilde{\mathcal{Y}}_c) = (\prod_{\mathcal{X}_{i,k}\in\mathcal{D}_\mathcal{X}} \mathbb{R}^{p^X_{i,k}}) \times (\prod_{\mathcal{Y}_{i,k}\in\mathcal{D}_\mathcal{Y}} \mathbb{R}^{p^Y_{i,k}})$, avec $\mathcal{D}_\mathcal{X} = \{\mathcal{X}_{i,k} | I(i,k) \text{ distincts}\}$ (idem pour les $\mathcal{Y}_i$)

Fonctions d'évaluation

On note dans la suite $\mathbf{X}_{i,c}$ les données projetées par injection canonique dans l'espace correspondant au type de données, bien définies pour tout i et tout c.

Definition

Soit H_c un espace de fonctions sur $(\tilde{\mathcal{X}}_c, \tilde{\mathcal{Y}}_c)$ à valeur dans $\mathbb{R}$ tel que pour tout $h \in H_c$:

- h est "suffisamment régulière"
- $q = \int_{(\tilde{\mathcal{X}}_c, \tilde{\mathcal{Y}}_c)} h$ est une fonction de qualification du fait urbain

Example

Pour une moyenne des lignes de $\mathbf{X}_{i,c}$, on aura $h(x) = x \cdot f_{i,c}$ avec $f_{i,c}$ densité de la distribution.
Pour un taux suivant une certaine condition, on aura $f_{i,c}\chi_{condition}$; etc
Pour les Y on aura en général un Dirac (pour les calculs aggrégés).

Pondération des critères

On donne des poids aux critères pour chaque réalisation :
Pour i, c et $h_c \in H_c$ donnés, $w_{i,c}$ est une combinaison (suivant [Wang et al., 2009]) de :

- poids objectif : importance locale du critère $w^L_{i,c} = \frac{\hat{q}_{i,c}}{\sum_c \hat{q}_{i,c}}$ où $\hat{q}_{i,c}$ est un estimateur de q_c pour les données $\mathbf{X}_{i,c}$
- poids subjectif : différentes méthodes (notations, ordre d'importance etc) revues dans [Wang et al., 2009] pour donner un poids subjectif normalisé (à voir lesquelles)

Application des méthodes d'approximation d'intégrales

Theorem

[Varet, 2010]
Avec $\mathbf{X}_{i,c} = (\vec{X}_{i,c,l})_{1 \leq l \leq n_{i,c}}$, $D_{i,c} = Discp_{\tilde{\mathcal{X}}_{c},\infty}(\mathbf{X}_{i,c})$ *la discrépance du nuage de points ([Niederreiter, 1972]) et* $h \in H_c$, *on a la majoration de l'erreur sur l'intégrale*

$$\left\| \int h_c - \frac{1}{n_{i,c}} \sum_{l} h_c(\vec{X}_{i,c,l}) \right\| \leq K \cdot |||h_c||| \cdot D_{i,c}$$

On a alors directement
$\left\| \int \sum w_{i,c} h_c - \frac{1}{n_{i,c}} \sum_l w_{i,c} h_c(\vec{X}_{i,c,l}) \right\| \leq K \sum_c |||h_c||| \cdot D_{i,c}$

Definition

En supposant les espaces d'indicateurs normés, on définit alors un ratio de robustesse pour comparer deux évaluations de deux réalisations du fait urbain

$$R_{i,i'} = \frac{\sum_c w_{i,c} \cdot D_{i,c}}{\sum_c w_{i',c} \cdot D_{i',c}}$$

En définissant une relation d'ordre sur les réalisations par rapport à la position du ratio à 1, on définit un ordre complet sur les évaluations.
On a bien pris en compte l'impact de la structure des données sur la robustesse de l'évaluation. Le calcul du ratio ne dépend que de celui de la discrépance, qui est faisable dans ces dimensions [Bundschuh and Zhu, 1993] : package R `DiceDesign`.

Conclusion

- Semble satisfaisant théoriquement.
- Evaluation structurelle dont la méthode ne dépend pas du choix des indicateurs (classe très large)
- Pour la suite : implémentation sur nos cas particuliers (commencé : forme des fonctions h_c).

References I

Bundschuh, P. and Zhu, Y. (1993).
A method for exact calculation of the discrepancy of low-dimensional finite point sets i.
In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 63, pages 115–133. Springer.

Mangin, D. and Panerai, P. (1999).
Projet urbain.
Parenthèses.

Niederreiter, H. (1972).
Discrepancy and convex programming.
Annali di matematica pura ed applicata, 93(1):89–97.

References II

Varet, S. (2010).
Développement de méthodes statistiques pour la prédiction d'un gabarit de signature infrarouge.
PhD thesis, Université Paul Sabatier-Toulouse III.

Wang, J.-J., Jing, Y.-Y., Zhang, C.-F., and Zhao, J.-H. (2009).
Review on multi-criteria decision analysis aid in sustainable energy decision-making.
Renewable and Sustainable Energy Reviews, 13(9):2263–2278.

Questions

?