

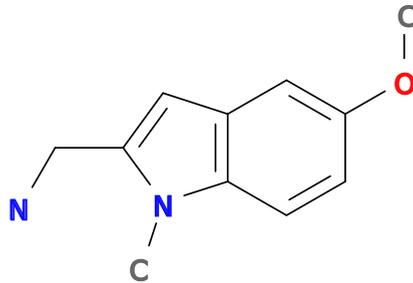
Apprentissage profond pour l'approximation de la distance d'édition entre graphes

G. Renton

Université de Rouen

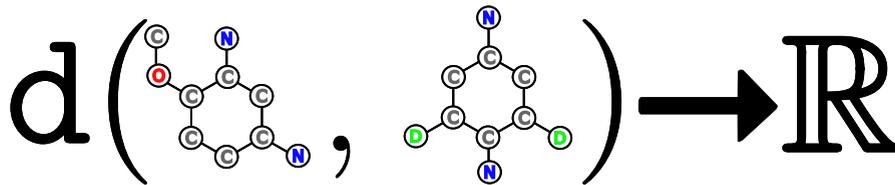
April 19, 2018

Introduction



Problématique

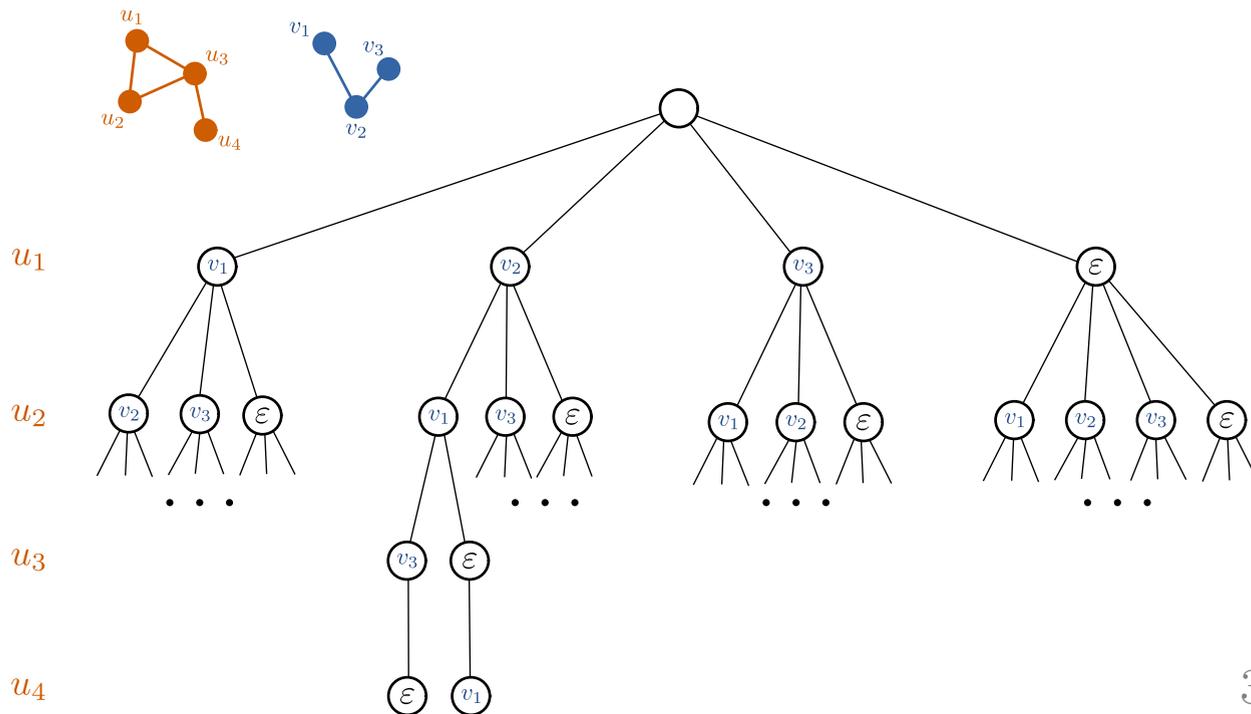
- Besoin d'une mesure de dissimilarité entre graphes.



Introduction

Distance d'édition entre graphes

- Coût de transformer g_1 en g_2 .

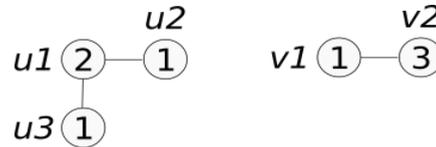


État de l'art

Appariement de graphe bipartite

- Appariement de nœuds deux à deux.
- Définition d'un coût d'appariement :

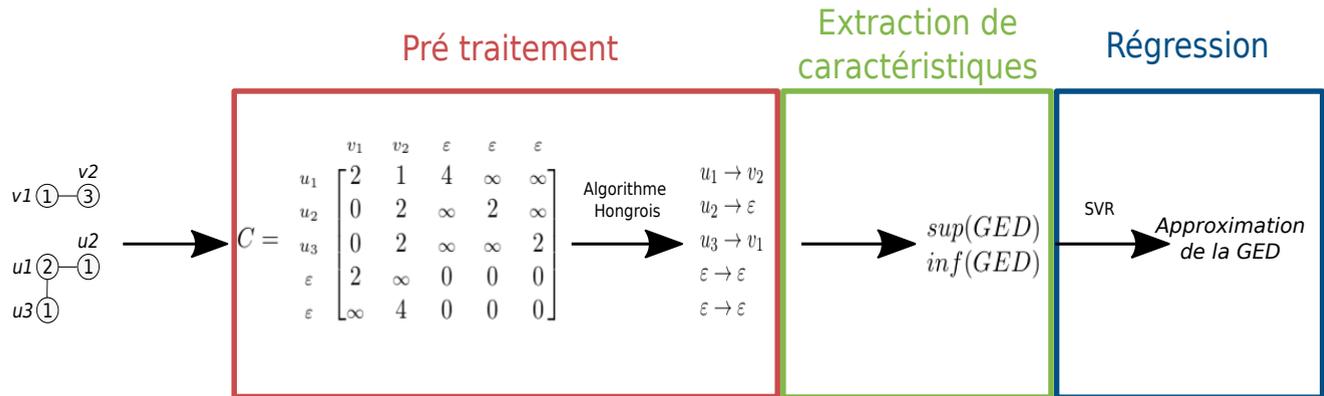
$$c(e_i) = c(v_i \rightarrow u_j) + \min \sum_{k=1}^{n+m} c(a_{ik} \rightarrow b_{j\varphi_k})$$



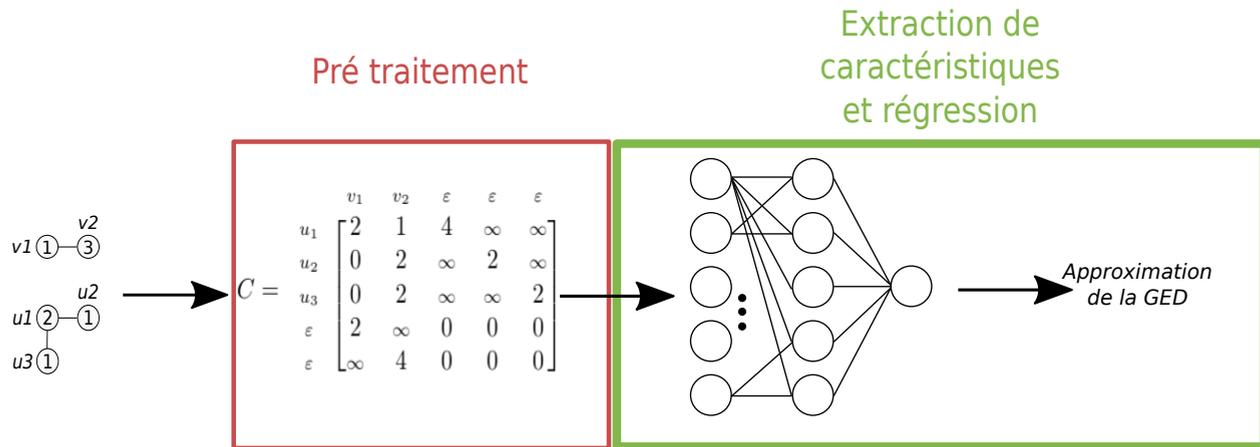
$$C = \begin{array}{c|cc|ccc} & v_1 & v_2 & \varepsilon & \varepsilon & \varepsilon \\ \hline u_1 & \begin{bmatrix} 2 & 1 \\ 0 & 2 \\ 0 & 2 \end{bmatrix} & \begin{bmatrix} 4 & \infty & \infty \\ \infty & 2 & \infty \\ \infty & \infty & 2 \end{bmatrix} \\ \hline \varepsilon & \begin{bmatrix} 2 & \infty \\ \infty & 4 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \hline \varepsilon & & & & & \end{array}$$

État de l'art

Apprentissage



Méthode Proposition



Méthode

Entrée du réseau

Objectif : rendre moins sensible aux permutations

→ Permutation de C

→ Les solutions de l'algorithme Hongrois forment la diagonale.

$$C = \begin{array}{c} u_1 \\ u_2 \\ u_3 \\ \varepsilon \\ \varepsilon \end{array} \begin{array}{cc} v_1 & v_2 \\ \left[\begin{array}{ccccc} 2 & 1 & 4 & \infty & \infty \\ 0 & 2 & \infty & 2 & \infty \\ 0 & 2 & \infty & \infty & 2 \\ 2 & \infty & 0 & 0 & 0 \\ \infty & 4 & 0 & 0 & 0 \end{array} \right] \end{array}$$

$$C = \begin{array}{c} u_1 \\ u_2 \\ u_3 \\ \varepsilon \\ \varepsilon \end{array} \begin{array}{cc} v_2 & \varepsilon \\ \left[\begin{array}{ccccc} 1 & \infty & 2 & \infty & \infty \\ 2 & 2 & 0 & \infty & 2 \\ 2 & \infty & 0 & 2 & \infty \\ \infty & 0 & 2 & 0 & 0 \\ 4 & 0 & \infty & 0 & 0 \end{array} \right] \end{array}$$

Réseau convolutif

- Application de filtres appris sur une matrice.
- Filtres de petites tailles (3×3 , 5×5)
- Permet de travailler sur des matrices de tailles variables.
- Requier d'augmenter la taille des champs réceptifs.

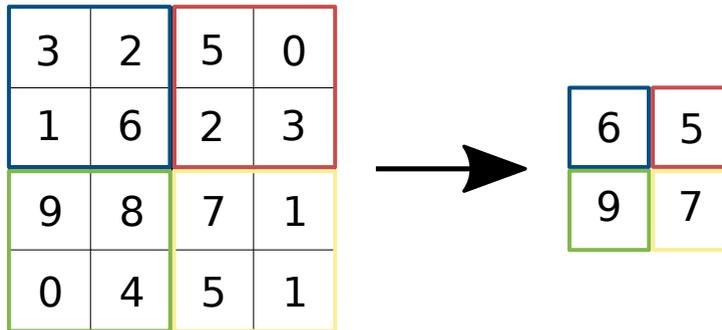
Deux stratégies :

- Le pooling
- La convolution dilatée

Méthode

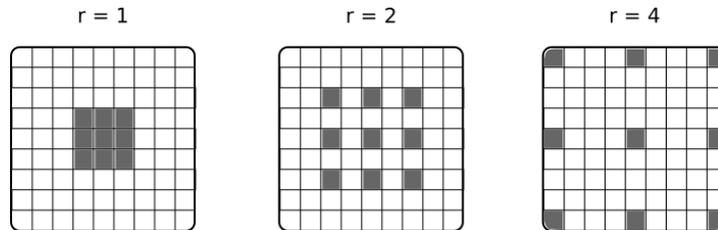
Pooling

- Réduction de la taille de la matrice.
- Agrégation des valeurs d'une fenêtre de petite taille.
- Représentation de l'ensemble des valeurs par une seule valeur (max, min, moyenne).



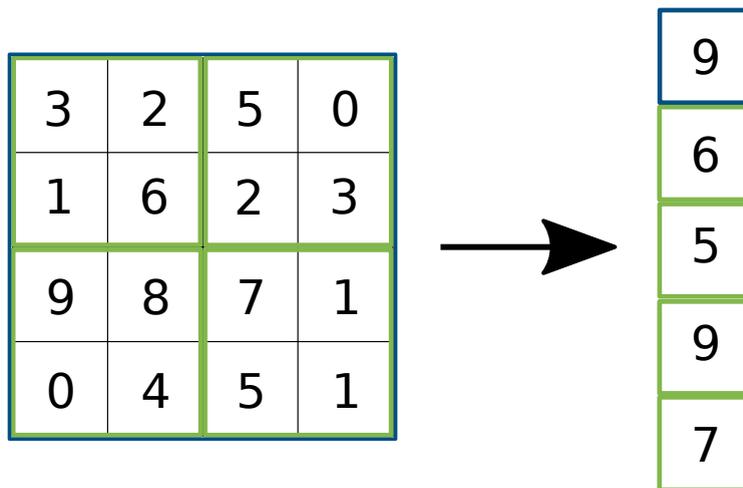
Convolution dilatée

- Augmentation de la taille des filtres.
- Taille augmentée avec des 0.
- Ajout d'un taux de dilatation.



Spatial Pyramid Pooling

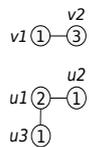
- Objectif : obtenir un vecteur de taille fixe.



Méthode

Modèle final

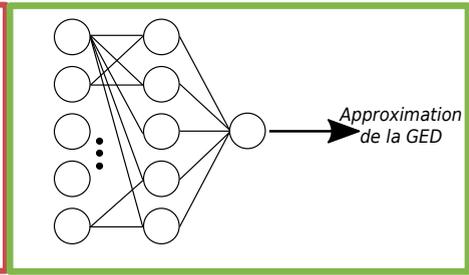
Pré traitement



$$C = \begin{array}{c} v_1 \quad v_2 \quad \varepsilon \quad \varepsilon \quad \varepsilon \\ \begin{array}{l} u_1 \\ u_2 \\ u_3 \\ \varepsilon \\ \varepsilon \end{array} \begin{bmatrix} 2 & 1 & 4 & \infty & \infty \\ 0 & 2 & \infty & 2 & \infty \\ 0 & 2 & \infty & \infty & 2 \\ 2 & \infty & 0 & 0 & 0 \\ \infty & 4 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$C = \begin{array}{c} v_2 \quad \varepsilon \quad v_1 \quad \varepsilon \quad \varepsilon \\ \begin{array}{l} u_1 \\ u_2 \\ u_3 \\ \varepsilon \\ \varepsilon \end{array} \begin{bmatrix} 1 & \infty & 2 & \infty & \infty \\ 2 & 2 & 0 & \infty & 2 \\ 2 & \infty & 0 & 2 & \infty \\ \infty & 0 & 2 & 0 & 0 \\ 4 & 0 & \infty & 0 & 0 \end{bmatrix} \end{array}$$

Extraction de caractéristiques et régression



Expériences

Protocole expérimental

- Expériences sur deux bases de graphes.
- Extraction de 1000 graphes.
- Pour chaque paire de graphes :
 - Calcul de la distance d'édition
 - Calcul de la matrice C
- Séparation en 3 bases : apprentissage, validation et test.

Évaluation

- Erreur relative moyenne :

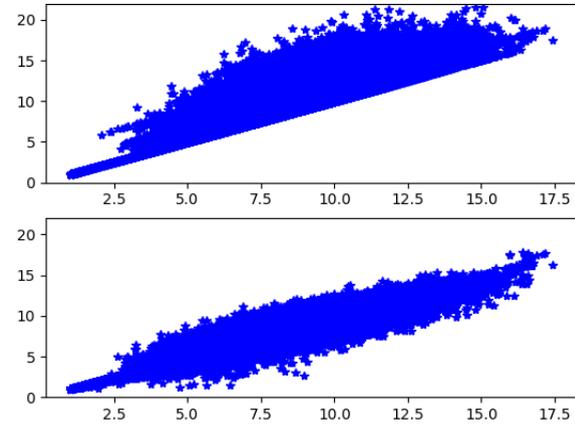
$$MRE = \frac{1}{N} \sum_{i=1}^N 100 * \frac{|d_{a_i} - d_{e_i}|}{d_{e_i}}$$

- Erreur quadratique moyenne :

$$MSE = \frac{1}{N} \sum_{i=1}^N (d_{a_i} - d_{e_i})^2$$

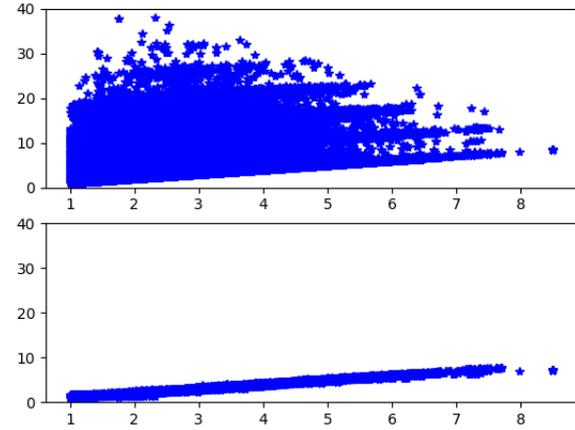
Expériences

Letter



	BP-GED	SVR	Approche proposée
MRE	13.11	7.76	4.2
MSE	3.07	0.86	0.32

Expériences Fingerprint



	BP GED	SVR	Approche proposée
MRE	68.39	7.56	1.15
MSE	16.15	0.10	0.006

Conclusion

Avantages

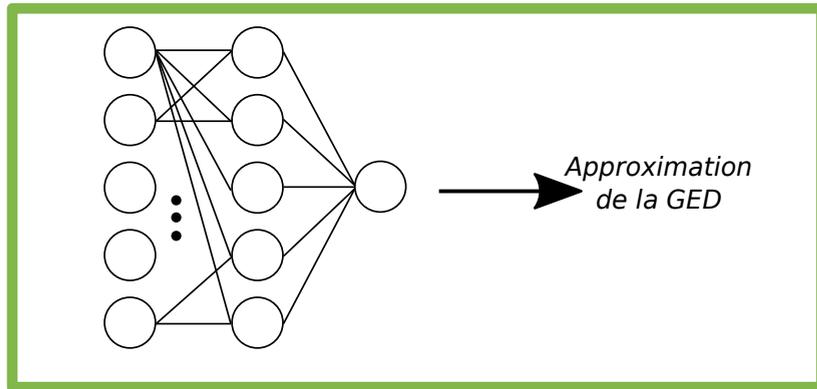
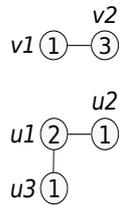
- Nette amélioration de la précision.
- Rapidité de décision.
- Complexité polynomiale avec le nombre de nœuds.

Inconvénients

- Nécessite une phase d'apprentissage.
- Apprentissage nécessite de connaître la vraie GED.

Perspectives

Extraction de
caractéristiques
et régression



*Approximation
de la GED*