

Vers un noyau de graphes efficace basé sur l'entropie

Aymen Ourdjini*, Abd Errahmane Kiouche**, Hamida Seba**

* Ecole nationale Supérieure d'Informatique (ESI) Oued Smar Alger Algérie
ga_ourdjini@esi.dz,
<https://www.esi.dz/>

** Université de Lyon, Université Lyon 1, LIRIS UMR 5205 F-69622 France
{abd-errahmane.kiouche, hamida.seba}@univ-lyon1.fr

1 Introduction

Les noyaux de graphes sont des algorithmes d'apprentissage automatique supervisé sur les graphes. Ils sont particulièrement appréciés pour leur efficacité en terme de précision. Cependant, la majorité des noyaux de graphes proposés dans la littérature ne sont pas assez rapides. Pour y remédier, nous proposons dans ce travail, un nouveau noyau de graphes basé sur le concept d'entropie. Notre méthode étend un noyau de graphes existant basé sur l'entropie de Rényi (Xu et al., 2021). Dans un graphe, l'entropie mesure la quantité d'information stockée dans un graphe. Ce qui revient à quantifier la complexité et le niveau d'organisation des caractéristiques structurelles de celui-ci. Le noyau qu'on propose améliore la précision du noyau de Xu et al. (2021) tout en réduisant son temps de calcul. Nos principales contributions sont :

- Nous avons remplacé l'entropie de Rényi par l'entropie de Von Neumann (Minello et al., 2019) pour tenir compte à la fois des relations de voisinage et de la distribution des degrés. Ce qui nous permet d'améliorer la précision du noyau ;
- Notre noyau tient également compte des informations auxiliaires portées par les nœuds comme les attributs ou les étiquettes (types) ;
- Nous proposons une nouvelle stratégie de calcul des scores de similarité entre deux graphes qui a la particularité d'être plus rapide ;
- Notre noyau proposé est parmi les plus rapides en terme de temps d'exécution ;

2 Notre noyau de graphes

Dans ce travail, nous proposons un nouveau noyau de graphe basé sur l'entropie de Von Neumann (Minello et al., 2019). Contrairement à la méthode de Xu et al. (2021), notre noyau est applicable sur les graphes étiquetés ou attribués. Notre méthode prend en entrée deux graphes étiquetés ou attribués. La première étape consiste à extraire le sous-graphe induit du voisinage de chaque nœud dans les deux graphes. Ensuite, nous calculons le score d'entropie de chaque nœud dans les deux graphes en appliquant l'entropie de Von Neumann (Minello et al., 2019) et en tenant compte des types (étiquettes) ou des attributs des nœuds. La dernière

étape consiste à calculer le noyau entre les deux graphes en se basant sur les scores d'entropie des nœuds des deux graphes à comparer. Nous appliquons le noyau RBF (radial basis function) pour calculer le noyau final de similarité. Nous avons appelé notre noyau **EVEGK (Enhanced Von-Neumann Entropy Graph Kernel)**.

Nous avons évalué les performances de notre noyau de graphes sur plusieurs ensembles de données. Tous les ensembles de données sont publiquement accessibles (Kersting et al., 2016). Notre approche est comparée au noyau d'entropie de Rényi (Second order Rényi Entropy Graph Kernel) **SREGK** proposé par (Xu et al., 2021). Nous avons également comparé notre approche à plusieurs autres noyaux de graphes proposés dans la littérature. Ces noyaux sont (1) Shortest path kernel (**SP**), (2) Graph Hopper Kernel (**GH**), (3) Random Walk Kernel (**RW**), (4) Graphlet Sampling kernel (**GS**), (5) Neighborhood Hash Kernel (**NH**), (6) Weisfeiler-Lehman Optimal Assignment (**WL-OA**), et (7) Neighborhood Subgraph Pairwise Distance (**NSPD**). Dans nos expérimentations, nous utilisons la validation croisée (10-folds cross validation) en appliquant la classification par C-SVM pour calculer l'accuracy de la classification. La table 1 illustre les scores d'accuracy de classification de tous les noyaux considérés, sur les 6 ensembles de données étiquetés. Les résultats montrent que notre noyau est le plus rapide sur 5 ensembles de données. En terme d'accuracy de classification, notre noyau n'a pas obtenu les meilleures performances, mais il s'est montré compétitif par rapport aux autres noyaux. Nous pouvons observer que sur tous les ensembles de données sauf AIDS notre approche est plus performante que l'approche **SREGK**. Cela prouve l'utilité et l'efficacité des améliorations que nous avons proposées.

TAB. 1 – Accuracy de la classification (\pm écart-type)

Noyaux	MUTAG	ENZYMES	PTC_MR	PROTEINS	AIDS	MSRC_21C
SP	87.18(± 0.99)	62.03(± 0.74)	65.43(± 1.48)	76.74(± 0.56)	99.59(± 0.03)	85.72(± 0.67)
GH	83.34(± 0.29)	42.33(± 1.11)	58.62(± 1.5)	76.33(± 0.44)	99.42(± 0.04)	27.34(± 1.23)
RW	66.49(± 0.0)	16.67(± 0.0)	55.82(± 0.0)	<i>OUT – OF – MEM</i>	80.0(± 0.0)	13.88(± 0.0)
GS	76.97(± 0.38)	28.82(± 1.15)	57.18(± 0.48)	71.99(± 0.36)	80.23(± 0.04)	17.73(± 1.69)
NH	90.15(± 0.86)	58.58(± 0.53)	66.08(± 0.95)	75.77(± 0.26)	99.44(± 0.02)	63.39(± 1.16)
WL-OA	88.54(± 0.75)	59.25(± 1.08)	66.79(± 1.08)	76.11(± 0.37)	99.36(± 0.05)	80.78(± 0.86)
NSPD	85.97(± 1.04)	44.78(± 0.99)	61.11(± 1.23)	75.26(± 0.25)	97.7(± 0.14)	82.61(± 0.62)
SREGK	86.65(± 0.89)	44.53(± 0.9)	59.82(± 1.15)	71.52(± 0.21)	98.88(± 0.09)	15.84(± 0.94)
EVEGK	91.0(± 0.5)	58.43(± 0.49)	63.02(± 0.78)	73.46(± 0.63)	98.29(± 0.13)	68.06(± 0.79)

N. B. : Ce travail a été effectué dans le cadre du projet ANR Gladis ANR-20-CE39-0008.

Références

- Kersting, K., N. M. Kriege, C. Morris, P. Mutzel, et M. Neumann (2016). Benchmark data sets for graph kernels.
- Minello, G., L. Rossi, et A. Torsello (2019). On the von neumann entropy of graphs. *Journal of Complex Networks* 7(4), 491–514.
- Xu, L., L. Bai, X. Jiang, M. Tan, D. Zhang, et B. Luo (2021). Deep rényi entropy graph kernel. *Pattern Recognition* 111, 107668.