Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 7, n. 1, 2020.

Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Processamento de sinal em grafos: Teoria de amostragem e sua aplicação no aprendizado semi-supervisionado

Evaristo Calisto Nhassengo [1]
Instituto de Ciências Matemáticas e de Computação, USP, São Carlos, SP
Jose Alberto Cuminato [2]
Instituto de Ciências Matemáticas e de Computação, USP, São Carlos, SP
Luis Gustavo Nonato [3]
Instituto de Ciências Matemáticas e de Computação, USP, São Carlos, SP

1 Introdução

A teoria de amostragem de sinais em grafos estuda o problema da escolha do melhor subconjunto de vértices para a reconstrução perfeita do sinal através de sua amostra. Será generalizado o problema clássico da amostragem de sinas, considera-se o problema da amostragem de sinais definidos em grafos. No estudo de sinais em grafos a representação do domínio de frequência é dada através dos autovalores e autovetores do Laplaciano. Visto que a amostragem de sinais definidos num grafo G, pode ser visto como um aprendizado ativo em G. Este trabalho propõe estudar o efeito que a amostragem de sinais em grafos tem nos algoritmos de aprendizado semi-supervisionado. Estuda-se a performance dos algoritmos semi-supervisionados, usando como dados iniciais a amostra obtida pelo algoritmo proposto em [1] e compara-se a performance deste algoritmo com outros algoritmos ativos do estado da arte (*Locally Linear Reconstruction, error bound minimization approach of LLGC*). Estabelece-se a relação entre a teoria de processamento de sinais em grafos [2] e o aprendizado semi-supervisionado.

2 Definições Básicas

Neste trabalho considera-se funções definidas grafos simples, conexos e não direcionados. A representação do grafo será através do seu Laplaciano. Um grafo $G = (\mathcal{V}, E)$, com o conjunto de vértices $\mathcal{V} = \{v_0, v_1, ..., v_{n-1}\}$ e arestas $E \subseteq \mathcal{V} \times \mathcal{V}$. A **matriz de adjacência** de G é uma matriz $n \times n$ com $A_{ij} = a_{ij}$. Um **sinal f do grafo G** é o mapeamento $f : V \to \mathbb{R}$. O **Laplaciano** de um grafo é definido como $L = D - A$, D é uma matriz diagonal $d_{ii} = \sum_k A_{ik}$. Considera-se o **Laplaciano normalizado**, definido

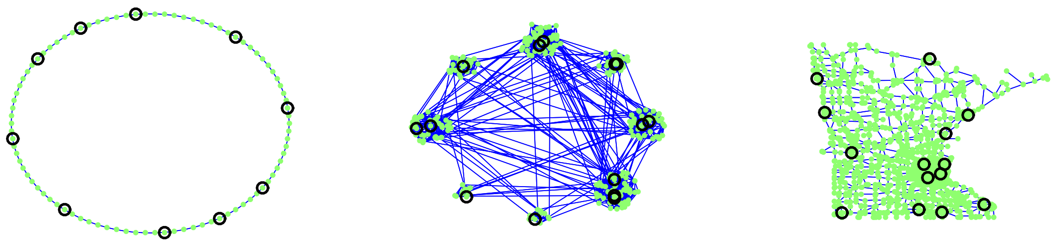
[1]evaristonhassengo717@gmail.com
[2]jacuminato@gmail.com
[3]lgustavo.nonato@gmail.com

como $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$. $\mathcal{L}$ é simétrica positiva semi-definida e tem o conjunto de autovalores reais $0 = \lambda_1 < \lambda_2 \leq ... \leq \lambda_n \leq 2$ e o conjunto de autovetores ortonormais $u_1, u_2, ..., u_n$. Define-se a transformada de Fourier (**GFT**) e a sua inversa (**iGFT**) como $\hat{f}(\lambda_\ell) = \sum_{i=1}^{n} f(v_i)u_\ell(v_i)$ e $f = \sum_{\ell=1}^{n} \hat{f}(\lambda_\ell)u_\ell$ respectivamente, onde u_ℓ são os autovetores e λ_ℓ são os autovalores de $\mathcal{L}$.

3 Problema, Proposta e Experimentos Preliminares

Nesta seção apresenta-se formalmente o problema da amostragem e a proposta de como proceder para melhorar a performance dos algoritmos semi-supervisionados. Considere o problema da extração de um conjunto de $\mathcal{S}$ coeficientes do sinal $x \in \mathbb{R}^n$ para produzir um sinal amostrado $x_{\mathcal{S}} \in \mathbb{R}^m$ $(n > m)$, onde $\mathcal{S} = (\mathcal{S}_0, ..., \mathcal{S}_{m-1})$ denota a sequência de índices amostrados e $\mathcal{S}_i \in \{0, 1, ..., n-1\}$, e então interpolar $x_{\mathcal{S}}$ para obter $x' \in \mathbb{C}^n$ que pode recuperar x aproximadamente.

Para melhoria da performance dos algoritmos semi-supervisionados propõe-se: Primeiramente recebe-se um conjunto de pontos $x \in \mathbb{R}^n$, em seguida constrói-se um grafo de similaridades (**KNN**) $G = (\mathcal{V}, A)$, extrai-se a amostra mais informativa $|\mathcal{S}| = l$, define-se o sinal na amostra ótima, propaga-se o sinal da amostra segundo algum algoritmo semi-supervisionado e finalmente estima-se o erro de classificação dos algoritmos semi-supervisionados e avalia-se as vantagens e desvantagens de usar os algoritmo de amostragem proposto em [1]. Nos experimentos preliminares foram utilizados 3 grafos para testar a performance do algoritmo de amostragem [1] e visualizar a amostra em grafos. A figura 1 mostra a performance deste algoritmo. Implicando uma boa reconstrução do sinal/rótulo a ser definido no grafo.



(a) Grafo Circular, $n = 90$, $|\mathcal{S}_{opt}| = 10$.

(b) Grafo de Comunidades $n = 256$ e $|\mathcal{S}_{opt}| = 16$.

(c) Grafo de Minesota, $|\mathcal{S}_{opt}| = 14$.

Figura 1: Amostras extraídas em grafos. Os vértices circundados pelas bolas pretas são os vértices amostrados e $\mathcal{S}_{opt}$ a amostra ótima segundo algoritmo de [1].

Referências

[1] A. Anis, A. Gadde and A. Ortega. Efficient Sampling Set Selection for Bandlimited Graph Signals Using Graph Spectral Proxies, *IEEE Trans. Signal Processing*, volume 64, 2016. DOI: 10.1109/TSP.2016.2546233

[2] A. Ortega, et all. Graph Signal Processing: Overview, Challenges and Applications, *Proceedings of the IEEE.*, volume 5, 2018. DOI: 10.1109/JPROC.2018.2820126.