Signal processing on simplicial complexes

`20221208. Thur.`

2106.07471.pdf (arxiv.org)

▼ 참고

- **Spectrogram classification**
- **Audio Deep Learning Made Simple: Sound Classification, Step-by-Step**
- **[Paper Review] Spectral-based Graph Convolutional Networks(GCN)**

검색어 : spectrogram classification, signal

Abstract

1. Higher order network : 복잡한 시스템을 갖는 multi-way relations, **다중연결 그래프**로 보겠다!
2. 최근 전염병 확산 과정, 의견 형성 등의 문제에서 다중연결성을 설명하려는 연구들 진행됨
3. 그 중에서도, 신호 데이터를 다중연결 그래프로 표현하고 GNN 적용할 것이다!
4. 기존의 시계열/이미지 관련 신호 처리 아이디어를 어떻게 그래프와 Simplicial complex로 확장?
    a. 기존 음성 신호 처리 연구 :
        - 시계열 데이터(1D)로 변환하여 분류하는 문제 : `accuracy value: 92.5128%`
        - 1D model + skip connection : `accuracy value: 93.1845%`
        - 2D spectrogram (STFT, short time fourier transform) : `accuracy value: 96.6614%`
        - 2D spec-skip : `accuracy value: 96.6021%`
5. 관련 개념
    a. Fourier analysis (푸리에 해석)
    b. signal denoising (신호 잡음 제거)
    c. signal interpolation (신호 보간) : 결측치(이상치) 관련
    d. nonlinear processing (비선형 처리)
6. Key Developments :
    a. hodge laplacian matrix (호지 라플라스 행렬)
    b. multi-relational operator (다중 관계 연산자) : simplicial cplx의 특별한 구조를 활용
    c. 그래프 신호 처리 문제에 대하여 라플라스 행렬의 개념/성질들을 합당하게 일반화

Introduction

1. system이 주어지면,
    a. fundamental entities of system ↔ nodes (즉, **시스템의 기본 요소는 node에 대응**)

b. their relations ↔ edges (**요소들 간의 관계는 edge에 대응**)

in a graph

2. 응용 : biology, social sciences, etc.
3. **그래프의 역할은 데이터나 프로세스의 유형에 영향을 받음**
4. (대충..) 크게 두 관점으로 나누자!
   a. **데이터를 네트워크 자체의 연결 패턴으로** 보자!
      - 목표 : 네트워크에서 주어진 high-dimensional 시그널의 이해
      - 이 관점은, 관계형 데이터를 모델링 할 때의 중심 아이디어
        - **관계형 데이터** : 네트워크에서 연결성이나 상호작용 같은 측정된 edge들에 대응되는 데이터 (**즉, edge들의 측정값을 다시 데이터node로 삼음**)
        - 예컨대, community detection(커뮤니티 감지), centrality analysis(중심성 분석) 또는 다양한 도구의 range 등을 통해 **통계적 연결고리(stochastic connections)에서 패턴을 찾아 시스템에 대해 학습**하는 걸 목표로 함

   b. 네트워크 위의 **dynamical process들을 연구**할 때, **네트워크는 임의**로 가져오되, **nodes(entities)는 고정**시키자!
      - 목표 : 그래프 구조를 활용하여, **nodes와 관련된 dynamics를 이해** (즉, 그래프로 주어지는 데이터 이해하기)
      - 문제 제기 : **그래프 노드로 주어지는 데이터와 프로세스를 이해하는 문제**는 “network dynamical system”의 맥락에서만 발생하는 건 아님 (이러한 유형의 데이터에 중점을 둔 또 다른 영역은 **GSP** → 이 논문에선 이걸 다룰 것임)
      - **GSP (graph signal prosecing)** : 네트워크의 노드로 주어지는 general signal 분석과 관련, **data (node) = signal + noise**
        - 이러한 **signal들은** 각 node 위에서의 dynamical 측정값으로부터 얻어진 **시계열 형태로 볼 수 있지만,**
        - 많은 signal들은 **data로부터 얻은 (다른 유형의) attribute data를 포함**
          - 다른 유형의 attribute data :
            - **node-covariates**
            - **node meta-data**
            - **node feature data**
        - GOAL of GSP
          - 개념 확장 : 기존의 신호 처리에서의 ‘**푸리에 변환**이나 **필터링 연산자들의 집합**’ 을 그래프 데이터로 잘 확장하자!

GSP (graph signal processing)

지난 10년 간… graph기반 데이터 분석 및 처리의 다양한 방법이 제공되고 꾸준히 관심 증가

1. (시스템의 그래프 기반 설명에 대한) **지금까지의 패러다임** : **multiple layers**나 **temporal dimensions**을 통해 **(아마도 학습 과정에서?) 잠재적으로 늘어나는 “heterogeneous”로 구성된 시스템과 데이터를 분석**하는 것이 주류
2. 복잡한 시스템의 특정 측면을 모델링할 때, **graph의 유용성**에 대한 조사와 연구도 비교적 최근

3. 문제점 : 복잡한 시스템에서는 다중연결이 흔하지만 **그래프는 둘 이상의 node 사이의 연결성을 encoding하지 않음**
    - Remark : 논문에서는 “multi-way interactions”이라 표현했는데, 이산수학에서 다중연결 그래프 아니냐고…
    - 예 : assemblies of neurons fire in unison (뭔소리야), 생화학 반응은 일반적으로 두 가지 이상의 화학 종을 포함하며, 그룹 간 상호작용은 사회적 맥락에서 흔한 현상임
    - 이러한 **polyadic interactions을 표현**하기 위해, **simplicial complexes**, **hypergraphs** 등을 포함하는 **higher-order relations을 모델링**하는 많은 프레임워크가 제안됨
    - **최근** : 이러한 프레임워크를 사용하여 polyadic relational data에 대한 (higher-order edge) **구조의 조직 원리(? 그냥 구조 자체를 연구했다는 말 아님?)를 분석하는 것이 주목**받음
    - **한계점** : 이렇게 복잡한 시스템의 다중연결 구조를 표현하고 분석하는 연구들이 많아졌지만, 아직 higher-order network **구조에 대한 dynamical process들의 연구는 부족**함
    - 그러나, higher-order 네트워크에서의 전염병 확산(epidemic spreading)이나 diffusion, opinion formation과 같은 작업들은 빠르게 발전 중
    - **문제 의식** : (위와 유사하게) **higher-order 네트워크 위에서의 신호 처리는 아직 연구 많이 부족 (거의 없음)**
4. GOAL of this paper :
    - higher-order network interactions에 대한 **프레임워크 모델링**으로서, **simplicial complexes에 초점을 맞춘 (액세스 가능하고 일관된 방식의) higher-order network 위에서의 신호 처리에 대한 구조와 동력학 프로세스** 제시
    - higher-order network 위에서의 특정 동력학적 프로세스 연구와의 연결성을 밝히고, 향후 연구 방법 제안

2장

1. 이산 신호 처리의 몇 가지 핵심 원칙 검토
2. 이산 신호 처리와 선형 동역학계 사이의 밀접한 관련성 강조

3장

1. **시계열 문제를 그래프 위에서 다뤄지는 신호 처리 문제로** 자연스럽게 일반화하는 방법을 설명하기 위하여, **동력계를 이용한 신호 처리의 이해 방식 사용**

    (참고 : Signal processing (time series analysis) for scientific data analysis with Python: Part 4 | by Nita Ghosh | Analytics Vidhya | Medium)
2. 관련 응용 예제 3가지 제시 (참고: Signal processing (scipy.signal) — SciPy v1.9.3 Manual)
    a. **signal smoothing** and **denoising** for graph signals

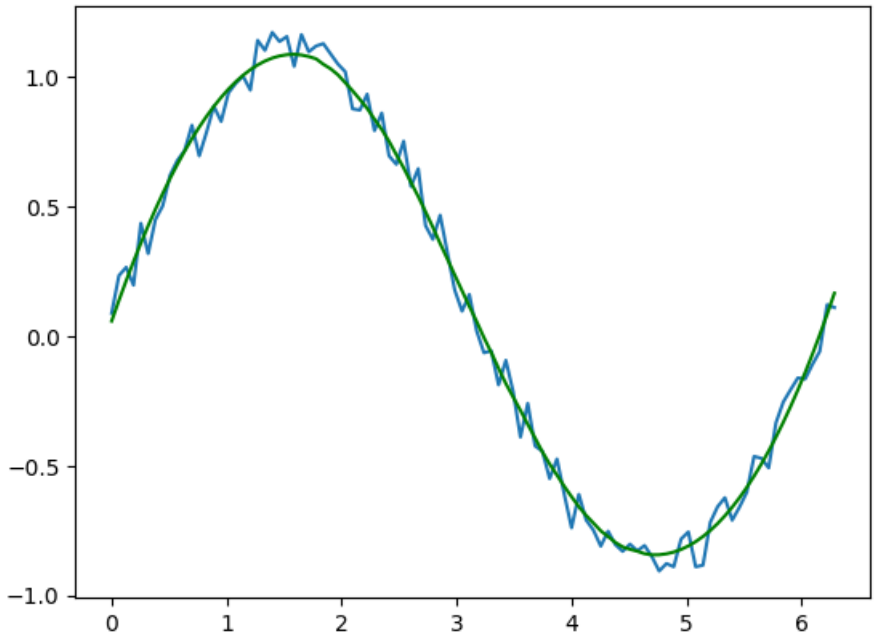
    ▼ smoothing

    1. `scipy.signal.savgol_filter()` 메서드 사용

```
import numpy as np
from scipy.signal import savgol_filter
import matplotlib.pyplot as plt

x = np.linspace(0,2*np.pi,100)
y = np.sin(x) + np.random.random(100) * 0.2
yhat = savgol_filter(y, 51, 3)
```

```
plt.plot(x, y)
plt.plot(x,yhat, color='green')
plt.show()
```



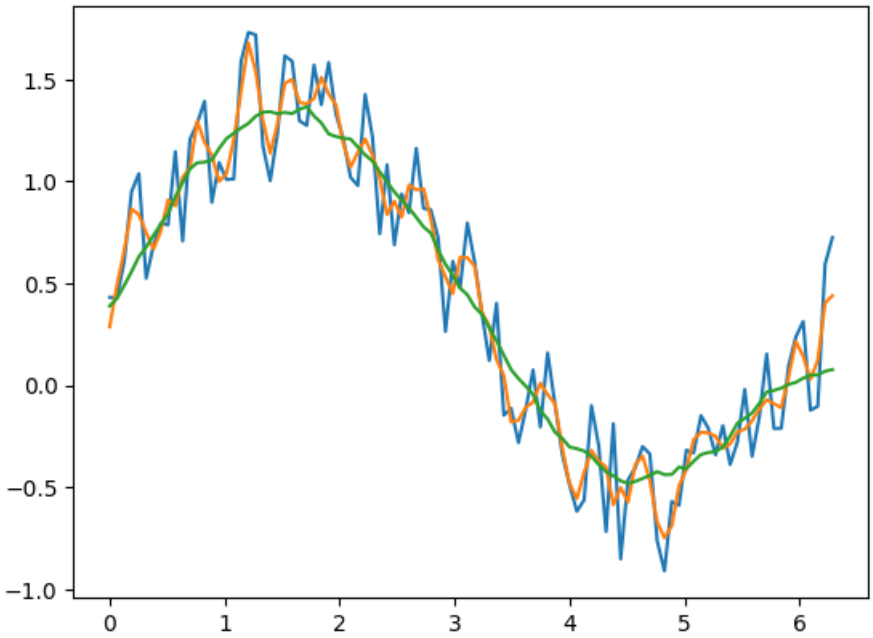
2. `numpy.convolve()` 메서드 사용

```
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(0,2*np.pi,100)
y = np.sin(x) + np.random.random(100) * 0.8

def smooth(y, box_pts):
    box = np.ones(box_pts)/box_pts
    y_smooth = np.convolve(y, box, mode='same')
    return y_smooth

plt.plot(x, y)
plt.plot(x, smooth(y,3))
plt.plot(x, smooth(y,19))
```



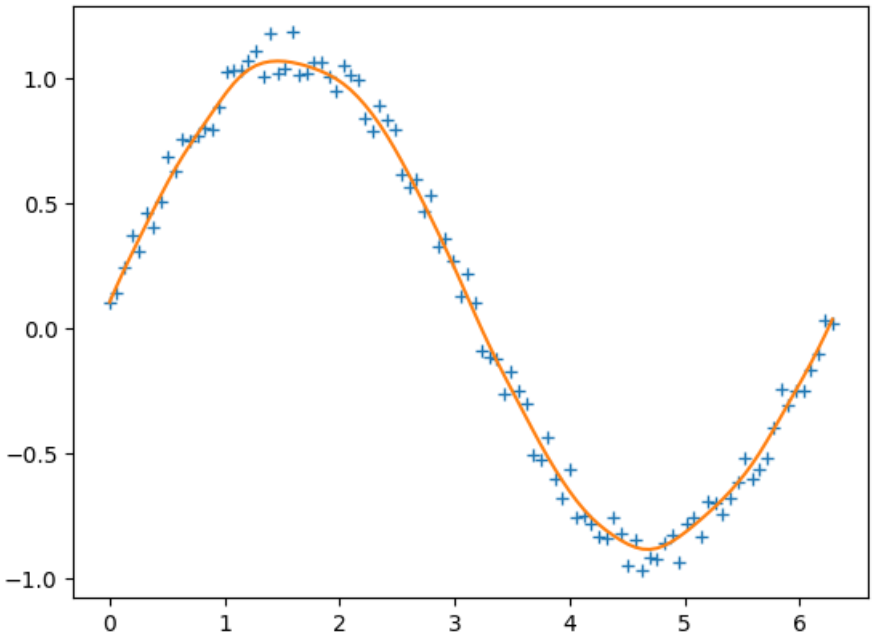
3. `statsmodels.nonparametric.kernel_regression` 의 `KernelReg()` 함수 사용 : 커널 회귀는 `y = g(X) + e` 의 조건부 평균 `E[y|X]` 계산

```
from statsmodels.nonparametric.kernel_regression import KernelReg
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(0,2*np.pi,100)
y = np.sin(x) + np.random.random(100) * 0.2

kr = KernelReg(y,x,'c')
plt.plot(x, y, '+')
y_pred, y_std = kr.fit(x)

plt.plot(x, y_pred)
plt.show()
```



▼ denoising 참고

- Fast Fourier Transform & Denoising | Kaggle
- Noise cancellation with Python and Fourier Transform | by Piero Paialunga | Towards Data Science
- Fft filters in Python/v3 (plotly.com)

b. **signal interpolation** on graphs

c. **nonlinear signal processing via GNN** (graph neural networks)

4장

1. 그래프 신호 처리로부터 higher-order network로 아이디어 확장
2. simplicial complexes (SC) 위에서의 신호 처리에 대한 중심 아이디어에 대한 대략적인 개요 언급
3. (3장-2) SC에 기반한 신호처리에서 (SC에 대한 graph laplacian의 일반화인) Hodge laplacian이 어떤 중요한 역할을 하는지 보일 것임
4. 그래프와 SC 위에서의 특정 동력학계 연구와의 관련성 언급
5. 향후 연구 방법 및 방향에 대한 간략한 논의

참고 : 오디오 데이터 증강 (Audio data augmentation)

1. **오디오 데이터 증강 - (PyTorch) 한국어 튜토리얼**
2. https://paintycode.tistory.com/30
3. https://kaen2891.tistory.com/74
4. **Seminar - 고려대학교 DMQA 연구실**