A Poisson-process AutoDecoder for X-ray Sources

Yanke Song
with Ashley Villar and Rafael Martínez-Galarza

Stats 300 Seminar

October 22, 2024

Contents

X-ray Sources

- X-ray surveys [1, 4, 2] produce massive X-ray data.
- The data contain event files of photon arrivals:

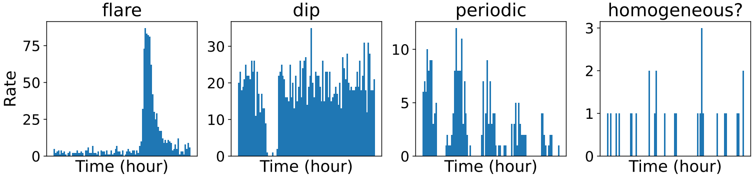
$$\{(t_i, e_i)\}_{i=1}^{n}$$

X-ray Sources

- X-ray surveys [1, 4, 2] produce massive X-ray data.
- The data contain event files of photon arrivals:

$$\{(t_i, e_i)\}_{i=1}^{n}$$

- Want to learn these sources automatically.
  - Source type classification
  - Anomaly detection



Previous Works

- Both supervised and unsupervised.

Previous Works

- Both supervised and unsupervised.
- One line of work: manual feature selection.
    - Requires domain knowledge.
    - May require time-consuming pipelines.

Previous works

Property name	Description
hard_hm	ACIS hard (2.0–7.0 keV) – medium (1.2–2.0 keV) energy band hardness ratio – basically the ratio between the hard and medium energy bands
hard_hs	ACIS hard (2.0–7.0 keV) – soft (0.5–1.2 keV) energy band hardness ratio – basically the ratio between the hard and soft energy bands
hard_ms	ACIS medium (1.2–2.0 keV) – soft (0.5–1.2 keV) energy band hardness ratio – basically the ratio between the medium and soft energy bands
bb_kt	Temperature (kT) of the best-fitting absorbed blackbody model spectrum to the source region aperture PI spectrum – temperature of the object estimated by a blackbody model.
powlaw_gamma	Photon index of the best fitting absorbed power-law model spectrum to the source region aperture
var_prob_*	Intra-observation Gregory–Loredo variability probability (highest value across all stacked observations) for each science energy band – variability probability in a single observation with Gregory–Loredo technique
var_ratio_*	The ratio of flux variability mean value to its standard deviation $\frac{\text{var\_mean\_*}}{\text{var\_sigma\_*}}$
var_newq_b	Proportion of the average of minimum and maximum count rates (i.e. data points in the light curve) during an observation relative to the mean count rate $\frac{\text{var\_max\_b} + \text{var\_min\_b}}{2 \cdot \text{var\_mean\_b}}$

Figure 1: Features selected in [3].

Previous Works

- Another line of work: deep learning
  - CNN, RNN, etc.
  - Requires reconstructed rate function: $\{r(\tau_i)\}_{i=1}^N$ where $\{\tau_i\}_{i=1}^N$ is a uniform grid.

Previous Works

- Another line of work: deep learning
  - CNN, RNN, etc.
  - Requires reconstructed rate function: $\{r(\tau_i)\}_{i=1}^N$ where $\{\tau_i\}_{i=1}^N$ is a uniform grid.
- Rate function reconstruction:
  - Naively bins the event file: artifacts due to Poisson arrival.

Previous Works

- Another line of work: deep learning
  - CNN, RNN, etc.
  - Requires reconstructed rate function: $\{r(\tau_i)\}_{i=1}^{N}$ where $\{\tau_i\}_{i=1}^{N}$ is a uniform grid.
- Rate function reconstruction:
  - Naively bins the event file: artifacts due to Poisson arrival.
  - Gregory Loredo algorithm: a full Bayesian approach
    - Consider all stepwise light curves up to a certain frequency.
    - Uniform Prior + Poisson likelihood.
    - Superimpose all proposals weighted by posterior.

Previous Works

- Another line of work: deep learning
  - CNN, RNN, etc.
  - Requires reconstructed rate function: $\{r(\tau_i)\}_{i=1}^{N}$ where $\{\tau_i\}_{i=1}^{N}$ is a uniform grid.
- Rate function reconstruction:
  - Naively bins the event file: artifacts due to Poisson arrival.
  - Gregory Loredo algorithm: a full Bayesian approach
    - Consider all stepwise light curves up to a certain frequency.
    - Uniform Prior + Poisson likelihood.
    - Superimpose all proposals weighted by posterior.
  - Drawbacks of GL:
    - Resolution limited due to computational complexity.
    - Only reconstructs rate function. Need separate pipeline for learning.

Previous Works

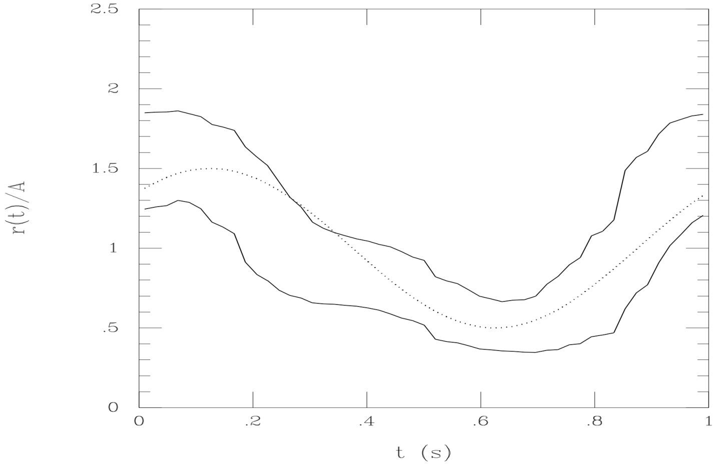


Figure 2: Reconstruction by GL algorithm.

Contribution

A learning pipeline that

Contribution

A learning pipeline that

- Is fully unsupervised

Contribution

A learning pipeline that

- Is fully unsupervised
- Respects the Poisson nature

Contribution

A learning pipeline that

- Is fully unsupervised
- Respects the Poisson nature
- Has adaptive resolution

Contribution

A learning pipeline that

- Is fully unsupervised
- Respects the Poisson nature
- Has adaptive resolution
- Is end-to-end: rate function reconstruction + representation learning

Loglikelihood

- Assume no energy marking for now: $\{t_i\}_{i=1}^{n}$

Loglikelihood

- Assume no energy marking for now: $\{t_i\}_{i=1}^{n}$
- Likelihood for a candidate rate function r:

$$\text{likelihood}(t_1, ..., t_n; r) = \left(\prod_{i=1}^{n} r(t_i)\right) \exp\left(-\int_0^T r(t)dt\right).$$

Loglikelihood

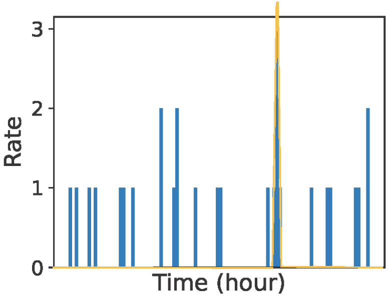
- Assume no energy marking for now: $\{t_i\}_{i=1}^n$
- Likelihood for a candidate rate function r:

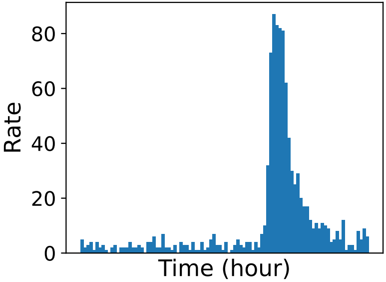
$$\text{likelihood}(t_1, ..., t_n; r) = \left(\prod_{i=1}^{n} r(t_i)\right) \exp\left(-\int_0^T r(t)dt\right).$$

- Use negative log likelihood as the loss function.

Regularization

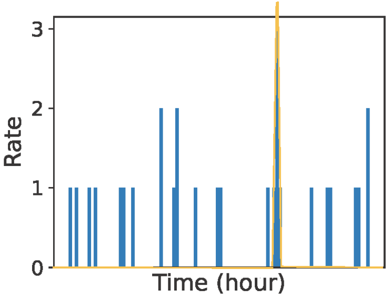
- What's the problem with likelihood only?

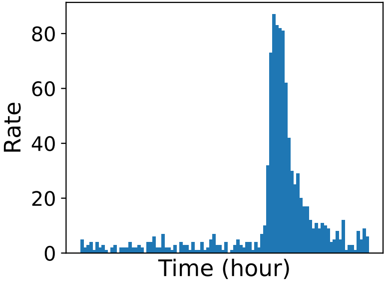




Regularization

- What's the problem with likelihood only?

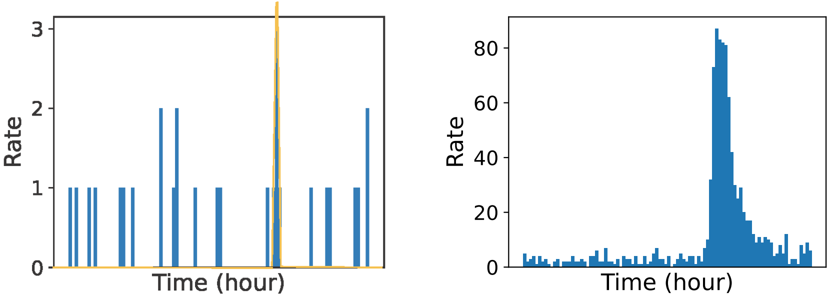




- Need to add "smoothness" regularization

Regularization

- What's the problem with likelihood only?

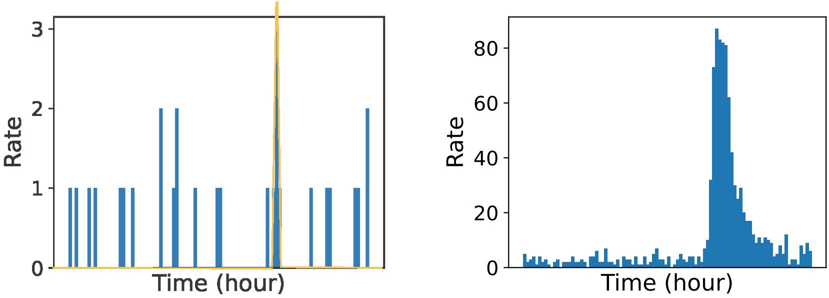


- Need to add "smoothness" regularization
- Total variation penalty:

$$\mathrm{TV}(r; \tau_1, ..., \tau_N) = \frac{1}{N-1} \sum_{i=1}^{N-1} |r(\tau_i) - r(\tau_{i+1})|$$

Regularization

- What's the problem with likelihood only?



- Need to add "smoothness" regularization
- Total variation penalty:

$$\mathrm{TV}(r;\tau_1,\ldots,\tau_N)=\frac{1}{N-1}\sum_{i=1}^{N-1}\left|r(\tau_i)-r(\tau_{i+1})\right|$$

- Two TV to guarantee enough coverage.

Neural representation

- Naive idea: parametrized basis function
  - Limited resolution
  - Need domain knowledge
  - Optimization

Neural representation

- Naive idea: parametrized basis function
  - Limited resolution
  - Need domain knowledge
  - Optimization
- Neural representation: approximate r with neural network r_ϕ.
- Infinite resolution: any input t, $r_\phi(t)$ yields the rate.

Neural representation

- Naive idea: parametrized basis function
  - Limited resolution
  - Need domain knowledge
  - Optimization
- Neural representation: approximate r with neural network r_ϕ.
- Infinite resolution: any input t, $r_\phi(t)$ yields the rate.
- Use a ResNet with ReLU activation
  - Not efficient in learning high frequencies.

Neural representation

- Naive idea: parametrized basis function
  - Limited resolution
  - Need domain knowledge
  - Optimization
- Neural representation: approximate r with neural network r_ϕ.
- Infinite resolution: any input t, $r_\phi(t)$ yields the rate.
- Use a ResNet with ReLU activation
  - Not efficient in learning high frequencies.
- Positional encoding:

$$\gamma(t) = [\bar{t}, \sin(2^0\pi\bar{t}), \cos(2^0\pi\bar{t}), ..., \sin\left(2^{L-1}\pi\bar{t}\right), \cos\left(2^{L-1}\pi\bar{t}\right)]. \quad (1)$$

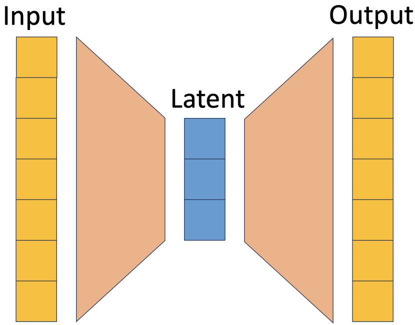
  where $\bar{t} = t/T$.
- Input $\gamma(t)$ to the network: $r_\phi(\gamma(t))$.

Representation learning

- Rate function reconstruction is complete. Where is the representation?

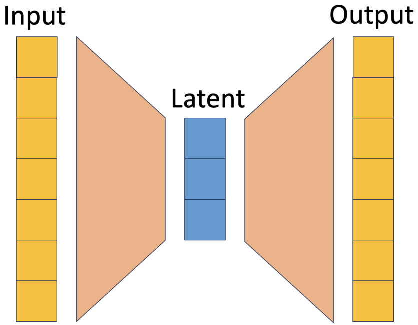
Representation learning

- Rate function reconstruction is complete. Where is the representation?
- (Unsupervised) representation learning for images: AutoEncoders



Representation learning

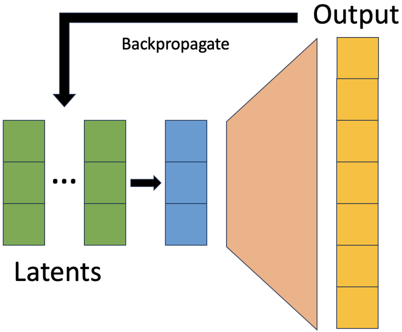
- Rate function reconstruction is complete. Where is the representation?
- (Unsupervised) representation learning for images: AutoEncoders



- What's the problem on event files?
  - Input has variable length.
  - Extremely low SNR
  - High variance in information throughput

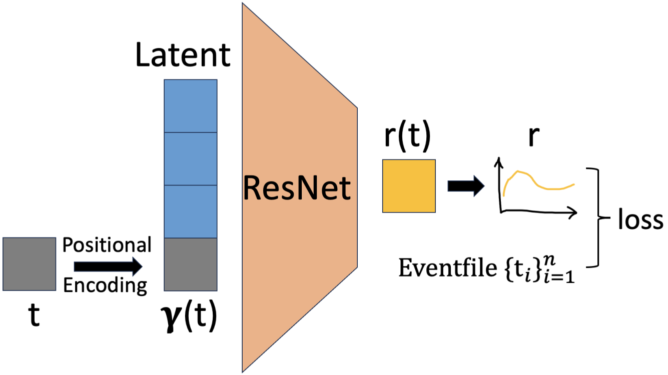
Autodecoders

- Autodecoder: no encoder!



- Directly "prepare" latent representations.
- Learn them together with the neural net.
- At test time: optimize the new latent.

PPAD


Latent
ResNet
r(t)
r
loss
Positional
Encoding
t
γ(t)
Eventfile $\{t_i\}_{i=1}^{n}$

PPAD

- $j = 1, ..., M$ event files, $k = 1, ..., K$ energy bins, $i = 1, ..., n_{j,k}$ events.

$$\mathcal{L}_{\text{total}}(\phi; \{z_j\}_{j=1}^{M}) = \sum_{j=1}^{M} \left(\sum_{k=1}^{K} \left(\mathcal{L}_{\text{neg-loglikelihood}}^{(j,k)} + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}}^{(j,k)} \right) + \lambda_{\text{latent}} \mathcal{L}_{\text{latent}}^{(j)} \right)$$

$$\mathcal{L}_{\text{neg-loglikelihood}}^{(j,k)} = -\sum_{i=1}^{n_{j,k}} \log r_{\phi}^{(k)}(\gamma(t_{i,k}); z^{(j)}) + \int_0^T r_{\phi}^{(k)}(\gamma(t); z^{(j)}) dt,$$

$$\mathcal{L}_{\text{TV}}^{(j,k)} = \left[\frac{1}{N-1} \sum_{i=1}^{N-1} |r_{\phi}^{(k)}(\gamma(\tau_i); z^{(j)}) - r_{\phi}^{(k)}(\gamma(\tau_{i+1}); z^{(j)})| \right.$$

$$\left. + \frac{1}{n-1} \sum_{i=1}^{n-1} |r_{\phi}^{(k)}(\gamma(t_i); z^{(j)}) - r_{\phi}^{(k)}(\gamma(t_i); z^{(j)})| \right],$$

$$\mathcal{L}_{\text{latent}}^{(j,k)} = \|z^{(j)}\|_2^2,$$

PPAD

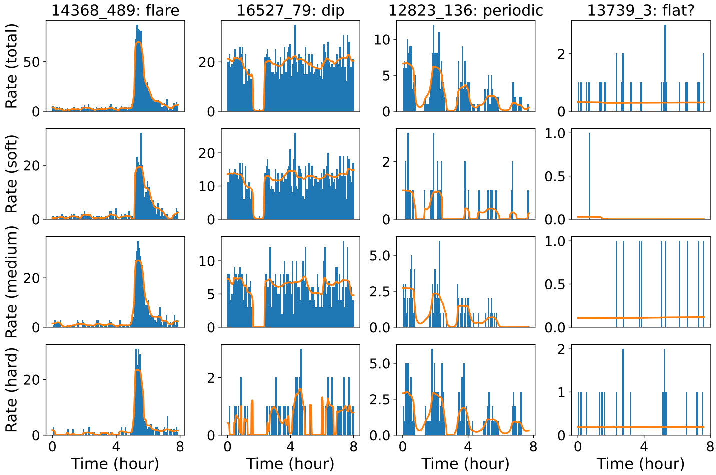
$$\textbf{Training}: \hat{\phi}, \{\hat{z}^{(j)}\}_{j=1}^{M} := \arg\min_{\phi; \{z_j\}_{j=1}^{M}} \mathcal{L}_{\text{total}}(\phi; \{z^{(j)}\}_{j=1}^{M}). \quad (2)$$

$$\textbf{Inference}: \hat{z} := \arg\min_{z} \mathcal{L}_{\text{total}}(\hat{\phi}; z). \quad (3)$$

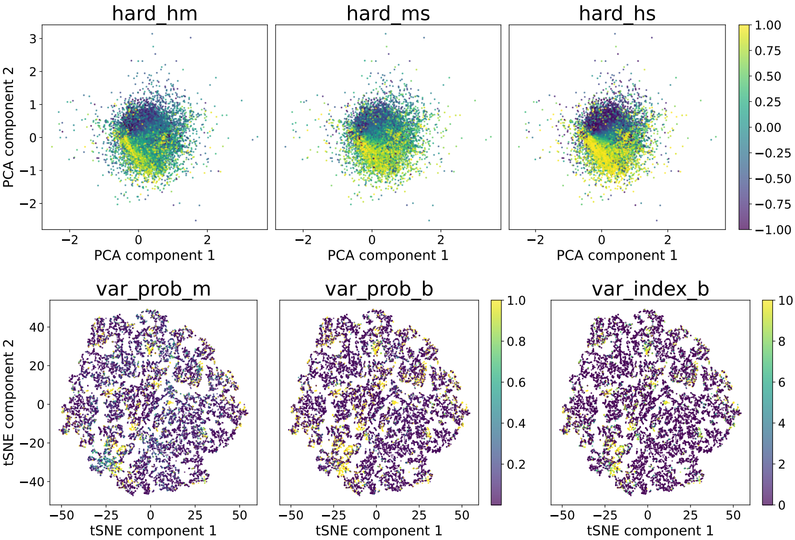
Data

- $\sim 10^5$ event files from the Chandra Source Catalog [1]
- Truncated to 8 hours
- Energy bins:
  - Soft: 0.5-1.2kV
  - Medium: 1.2-2kV
  - Hard: 2-7kV

Rate function reconstruction



Latent space

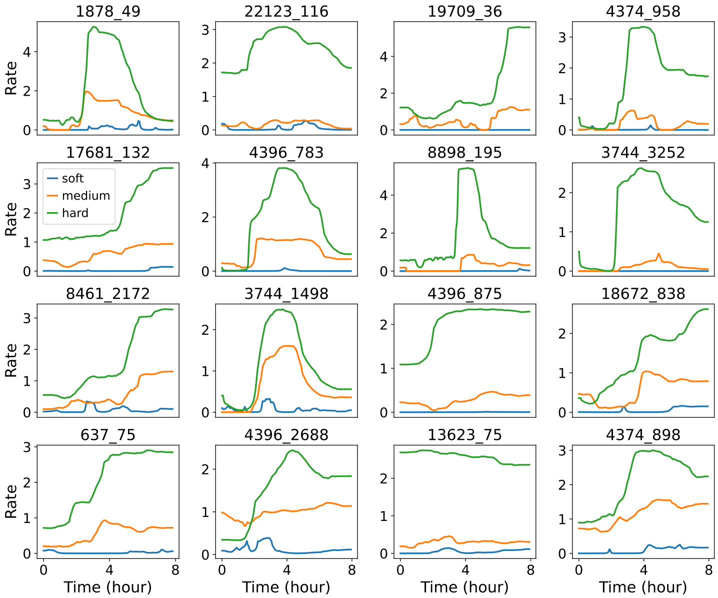


Prediction Results

Regression Traget	MSE	R^2
hard_ms	0.02	0.87
hard_hm	0.01	0.88
hard_hs	0.02	0.93
Classification Target	**Accuracy**	**F1 Score**
var_index_b > 5?	0.92	0.63
source type	0.62	0.25
YSO vs AGN	0.75	0.70

Table 1: Regression/classification performance using learned latent features. All models use a random forest with 100 trees and default hyperparameters Train-test split is $0.8 - 0.2$ without validation set. SMOTE is applied in classification case to resolve class imbalance.

Anomaly detection



Future Works

- Trade-off between reconstruction and representation.
- Allows sampling and UQ: variational autodecoders.
- Autoencoders.
- Invariance w.r.t. phase, total rate, etc.

Thank you!

References I

[1] Ian N. Evans et al. "The Chandra Source Catalog Release 2 Series". In: *arXiv e-prints*, arXiv:2407.10799 (July 2024), arXiv:2407.10799. DOI: 10.48550/arXiv.2407.10799. arXiv: 2407.10799 [astro-ph.HE].

[2] A. Merloni et al. "The SRG/eROSITA all-sky survey. First X-ray catalogues and data release of the western Galactic hemisphere". In: 682, A34 (Feb. 2024), A34. DOI: 10.1051/0004-6361/202347165. arXiv: 2401.17274 [astro-ph.HE].

[3] Víctor Samuel Pérez-Díaz et al. "Unsupervised machine learning for the classification of astrophysical X-ray sources". In: *Monthly Notices of the Royal Astronomical Society* 528.3 (2024), pp. 4852–4871.

[4] N. A. Webb et al. "The XMM-Newton serendipitous survey. IX. The fourth XMM-Newton serendipitous source catalogue". In: 641, A136 (Sept. 2020), A136. DOI: 10.1051/0004-6361/201937353. arXiv: 2007.02899 [astro-ph.HE].