3

[] O. Deiss. *Efficient labeling technique and interpretable deep neural network for the classification of seizures using continuous electroencephalograms*. PhD thesis, Georgia Institute of Technology, 2018.

[] K. Saab, J. Dunnmon, C. Ré, D. Rubin, and C. Lee-Messer. Weak supervision as an efficient approach for automated seizure detection in electroencephalography. *NPJ digital medicine*, 3(1):1–12, 2020.

[] Y. Yang, N. D. Truong, J. K. Eshraghian, A. Nikpour, and O. Kavehei. Weak self-supervised learning for seizure forecasting: a feasibility study. *Royal Society Open Science*, 9 (8):220374, 2022.

Weakly Supervised Bayesian Estimation of Seizure Likelihood

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. Prepare markers and eraser, erase the board.
2. Reset the timer.
3. Thank the panel for attending.

Outline

1. Introduction
   - Motivation
   - Prior work
   - Main contributions
2. Research Problem
3. Research Plan
   - Pipeline
   - Methods
   - Data
4. Empirical Results
5. Conclusion

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. Introduction
   - Motivation
   - Prior work
   - Main contributions
2. Research Problem
3. Research Plan
4. Empirical Results
5. Conclusion

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Background

Figure 1: 18th century iconograph of an epilepsy seizure. Etched by J. Duplessi-Bertaux. Credit: Wellcome Library, London. Licensed under the Creative Commons Attribution 4.0 International license.

1. `https://commons.wikimedia.org/wiki/File:A_group_of_people_standing_around_a_man_having_an_epileptic_Wellcome_L0005934.jpg`
2. In this iconograph from the 18th century, we see a group of people standing around a man having an epileptic fit.
3. Studies have shown that the biggest impediment to patients with epilepsy is the uncertainty and surprise factor of the seizures.

Background

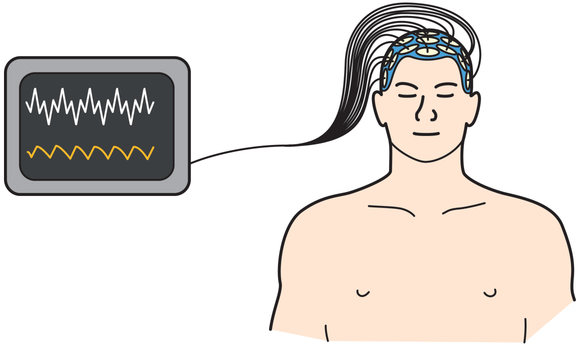


Figure 1: Electroencephalography (EEG) art.
From: The Clear Communication People. Licensed under: CC BY-NC-ND 2.0.

1. `https://commons.wikimedia.org/wiki/File:A_group_of_people_standing_around_a_man_having_an_epileptic_Wellcome_L0005934.jpg`
2. In this iconograph from the 18th century, we see a group of people standing around a man having an epileptic fit.
3. Studies have shown that the biggest impediment to patients with epilepsy is the uncertainty and surprise factor of the seizures.
4. EEG is a form of neuroimagery with high temporal accuracy. Seen here is a scalp eeg, based on a wearable cap which holds electrodes in contact with the head's surface. The electric potentials induced by the brain are recorded by a digital sampler for data analysis.

Background & Motivation

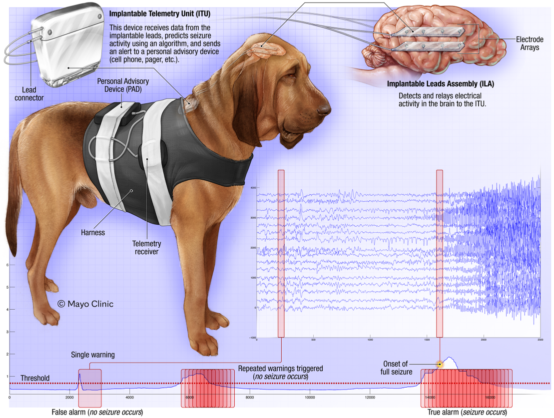
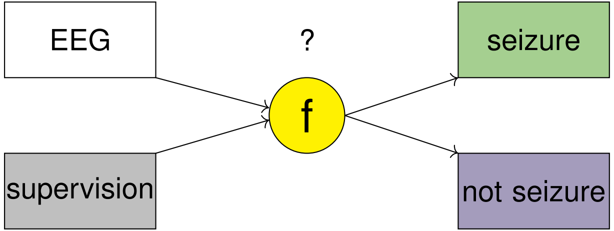


Figure 2: Canine Epilepsy Dataset. From: Coles et al. Feasibility study of a caregiver seizure alert system in canine epilepsy. Epilepsy Res. 2013 Oct; 106(3):456-60 Epub 2013. Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

1. In a seminal work by a mixed team of engineers and veterinarians, a seizure advisory system was evaluated in canines with naturally occuring epilepsy.
2. An intracranial EEG system implanted in a dog's brain was wirelessly transmitting the recordings to an external device.
3. 475 days of data were made available online.

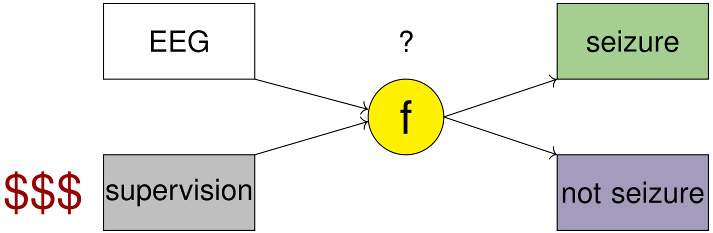
Background & Motivation



 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. For that reason, an active area of research since the 70's has been to predict epileptic seizures using EEGs.
2. A common paradigm which has been extensively studied, is supervised machine learning from labeled EEG segments.
3. But, this paradigm suffers from two problems.

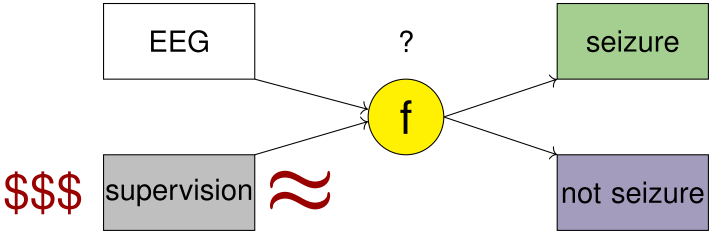
Background & Motivation



1. Expensive to produce.

1. For that reason, an active area of research since the 70's has been to predict epileptic seizures using EEGs.
2. A common paradigm which has been extensively studied, is supervised machine learning from labeled EEG segments.
3. But, this paradigm suffers from two problems.
4. First, labels are expensive to produce.

Background & Motivation

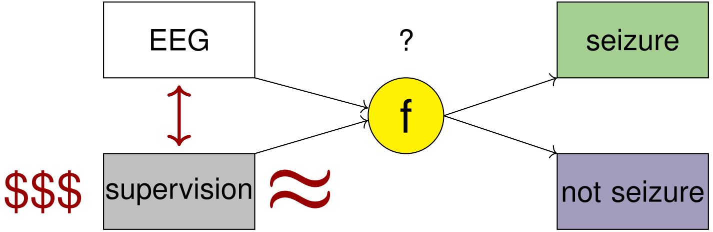


1. Expensive to produce.
2. Disagreements among experts.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. For that reason, an active area of research since the 70's has been to predict epileptic seizures using EEGs.
2. A common paradigm which has been extensively studied, is supervised machine learning from labeled EEG segments.
3. But, this paradigm suffers from two problems.
4. First, labels are expensive to produce.
5. second, EEG signal dynamics are more complex than the simplistic classes that experts have so far defined. Studies show that the labels are inexact, in the sense that disagreements between expert labelers are common in practice.

Background & Motivation



1. Expensive to produce.
2. Disagreements among experts.
3. Variables are tightly coupled.

1. For that reason, an active area of research since the 70's has been to predict epileptic seizures using EEGs.
2. A common paradigm which has been extensively studied, is supervised machine learning from labeled EEG segments.
3. But, this paradigm suffers from two problems.
4. First, labels are expensive to produce.
5. second, EEG signal dynamics are more complex than the simplistic classes that experts have so far defined. Studies show that the labels are inexact, in the sense that disagreements between expert labelers are common in practice.
6. Third, the annotations are tightly coupled to the EEG.
7. In other words: untaped seizures. The annotation, which could be collected in a handwritten seizure diary, are useless to the supervised learning paradigm. These types of models can't learn from annotations without EEGs.
8. Thus, partial or incomplete data points are completely ignored.

Prior work

1. Deiss [2018] introduces a co-learning process for improving label quality through active learning. However, the underlying deep-learning model requires a large labeled dataset.
2. Saab et al. [2020] cut costs by using weak annotations that are already provided as part of the standard clinical workflow. However, the model can not learn from seizure timestamps unless EEG was recorded at the event time, or vice versa.
3. Yang et al. [2022] show that weak self-supervised learning for seizure forecasting is feasible, *provided that* a pretrained detection model is available. Hence, their work depends on a supervised model which was trained on large labeled datasets.

Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. (1) Saab, Khaled, et al. "Weak supervision as an efficient approach for automated seizure detection in electroencephalography." NPJ digital medicine 3.1 (2020): 1-12.
2. (2) Deiss, Olivier. Efficient labeling technique and interpretable deep neural network for the classification of seizures using continuous electroencephalograms. Diss. Georgia Institute of Technology, 2018.
3. (3-prework) Yang, Yikai, et al. "Continental generalization of a human-in-the-loop AI system for clinical seizure recognition." Expert Systems with Applications 207 (2022): 118083.
4. (3) Yang, Yikai, et al. "Weak self-supervised learning for seizure forecasting: a feasibility study." Royal Society Open Science 9.8 (2022): 220374.

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.

1. (1)

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.
2. Demonstrating the approach on real-world data, achieving 0.88 AUC-ROC with zero labels on a seizure detection task.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. (1)
2. (2)

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.
2. Demonstrating the approach on real-world data, achieving 0.88 AUC-ROC with zero labels on a seizure detection task.
3. A weakly supervised version which is biased towards circadian rhythms is shown to improve detection in canines.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. (1)
2. (2)

Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

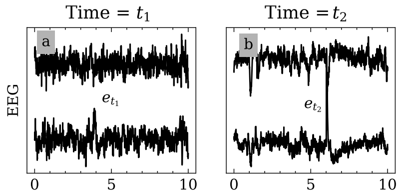


Figure 3: Samples from $E(t_1)$ and $E(t_2)$.

1. Defining a sample space and random variables.

Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

and:

2. $s_t \sim S(t) \in \Omega_S = \{0, 1\}$

a random Seizure variable.

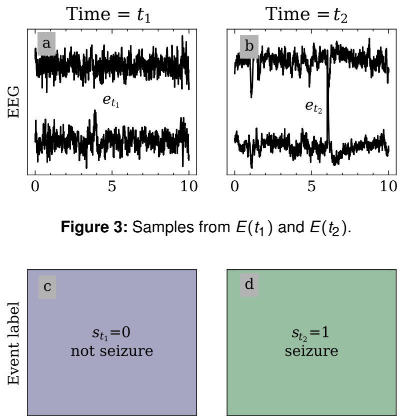


Figure 3: Samples from $E(t_1)$ and $E(t_2)$.

Figure 4: Samples from $S(t_1)$ and $S(t_2)$

Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. Defining a sample space and random variables.

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. There is a continuous EEG recording system sampling at f Hz. We are provided with a dataset D. For each time t, e_t is the observed EEG segment with c-channels of duration T, ending at time t. Also, a_t is a seizure annotation, provided by a board-approved expert. It is 1 if a seizure occurred in the interval ending at t (i.e. $[t - T, t]$) and 0 otherwise.

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

Solution

$$\underbrace{\mathbb{P}(S_t \mid E_t)}_{\text{probability of seizure given EEG}} = \frac{\mathbb{P}(S_t)\mathbb{P}(E_t \mid S_t)}{\mathbb{P}(E_t)} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Normalizing factor}}$$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. There is a continuous EEG recording system sampling at f Hz. We are provided with a dataset D. For each time t, e_t is the observed EEG segment with c-channels of duration T, ending at time t. Also, a_t is a seizure annotation, provided by a board-approved expert. It is 1 if a seizure occurred in the interval ending at t (i.e. $[t - T, t]$) and 0 otherwise.
2. The task of seizure forecasting with horizon τ is to find the likelihood of a seizure occurring between now and τ timesteps from now.

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

Solution

$$\underbrace{\mathbb{P}(S_t \mid E_t)}_{\text{probability of seizure given EEG}} = \frac{\mathbb{P}(S_t)\mathbb{P}(E_t \mid S_t)}{\mathbb{P}(E_t)} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Normalizing factor}}$$

Part two

How can we compute these values?

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. There is a continuous EEG recording system sampling at f Hz. We are provided with a dataset D. For each time t, e_t is the observed EEG segment with c-channels of duration T, ending at time t. Also, a_t is a seizure annotation, provided by a board-approved expert. It is 1 if a seizure occurred in the interval ending at t (i.e. $[t - T, t]$) and 0 otherwise.
2. The task of seizure forecasting with horizon τ is to find the likelihood of a seizure occurring between now and τ timesteps from now.
3. The question we ask is, how can we formulate the task of seizure forecasting as a Bayesian inference problem?

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Pipeline

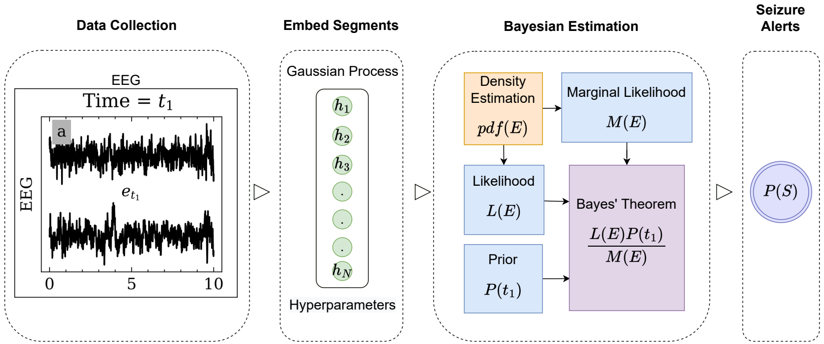


Figure 5: Bayesian Seizure Likelihood Estimation (BSLE) pipeline

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. Our method composes two stages: First, the EEG distribution is modeled and is used to assign novel sequences higher seizure likelihoods. Then, a priori knowledge based on the time covariate is added for supervised improvements.
2. Weakly Supervised Bayesian Estimation of Seizure Likelihood

Estimate seizure likelihood with Bayes' theorem

Goal

$$\mathbb{P}(S_t \mid E_t) = \frac{\mathbb{P}(E_t \mid S)\mathbb{P}(S_t)}{\mathbb{P}(E_t)} = \frac{L(E_t)P(t)}{M(E_t)}$$

1. $L(E_t)$ and $M(E_t)$ are computed with Monte Carlo Quantile Estimation. The likelihood is the quantile of the sample, trimmed at the HDR$_\alpha$*boundary*.P(t)*isapriorfunctionwhichreliesonlyonthetimecovariate*.
2. ! Write Equivalent goal on board

Estimate seizure likelihood with Bayes' theorem

Goal

$$\mathbb{P}(S_t \mid E_t) = \frac{\mathbb{P}(E_t \mid S)\mathbb{P}(S_t)}{\mathbb{P}(E_t)} = \frac{L(E_t)P(t)}{M(E_t)}$$

Equivalent goal

$$\mathbb{P}(S_t \mid E_t) = 1 - \mathbb{P}(\cancel{S_t} \mid E_t) = 1 - \frac{\mathbb{P}(E_t \mid \cancel{S})\mathbb{P}(\cancel{S_t})}{\mathbb{P}(E_t)} = 1 - \frac{L_{\cancel{S}}(E_t)P_{\cancel{S}}(t)}{M(E_t)}$$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. $L(E_t)$ and $M(E_t)$ are computed with Monte Carlo Quantile Estimation. The likelihood is the quantile of the sample, trimmed at the HDR$_\alpha$*boundary*.P(t)*isapriorfunctionwhichreliesonlyonthetimecovariate*.
2. ! Write Equivalent goal on board

Estimate seizure likelihood with Bayes' theorem

Goal

$$\mathbb{P}(S_t \mid E_t) = \frac{\mathbb{P}(E_t \mid S)\mathbb{P}(S_t)}{\mathbb{P}(E_t)} = \frac{L(E_t)P(t)}{M(E_t)}$$

Equivalent goal

$$\mathbb{P}(S_t \mid E_t) = 1-\mathbb{P}(\cancel{S}_t \mid E_t) = 1-\frac{\mathbb{P}(E_t \mid \cancel{S})\mathbb{P}(\cancel{S}_t)}{\mathbb{P}(E_t)} = 1-\frac{L_{\cancel{S}}(E_t)P_{\cancel{S}}(t)}{M(E_t)}$$

Method overview

1. $M(E_t)$ is the likelihood of observing E_t in general.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. $L(E_t)$ and $M(E_t)$ are computed with Monte Carlo Quantile Estimation. The likelihood is the quantile of the sample, trimmed at the HDR$_\alpha$*boundary*.P(t)*isapriorfunctionwhichreliesonlyonthetimecovariate*.
2. ! Write Equivalent goal on board

Estimate seizure likelihood with Bayes' theorem

Goal

$$\mathbb{P}(S_t \mid E_t) = \frac{\mathbb{P}(E_t \mid S)\mathbb{P}(S_t)}{\mathbb{P}(E_t)} = \frac{L(E_t)P(t)}{M(E_t)}$$

Equivalent goal

$$\mathbb{P}(S_t \mid E_t) = 1 - \mathbb{P}(\cancel{S}_t \mid E_t) = 1 - \frac{\mathbb{P}(E_t \mid \cancel{S})\mathbb{P}(\cancel{S}_t)}{\mathbb{P}(E_t)} = 1 - \frac{L_{\cancel{S}}(E_t)P_{\cancel{S}}(t)}{M(E_t)}$$

Method overview

1. $M(E_t)$ is the likelihood of observing E_t in general.
2. $L_{\cancel{S}}(E_t)$ is the likelihood of observing E_t provided $\cancel{S}$.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. $L(E_t)$ and $M(E_t)$ are computed with Monte Carlo Quantile Estimation. The likelihood is the quantile of the sample, trimmed at the HDR$_\alpha$*boundary*.P(t)*isapriorfunctionwhichreliesonlyonthetimecovariate*.
2. ! Write Equivalent goal on board

Estimate seizure likelihood with Bayes' theorem

Goal

$$\mathbb{P}(S_t \mid E_t) = \frac{\mathbb{P}(E_t \mid S)\mathbb{P}(S_t)}{\mathbb{P}(E_t)} = \frac{L(E_t)P(t)}{M(E_t)}$$

Equivalent goal

$$\mathbb{P}(S_t \mid E_t) = 1 - \mathbb{P}(\not{S}_t \mid E_t) = 1 - \frac{\mathbb{P}(E_t \mid \not{S})\mathbb{P}(\not{S}_t)}{\mathbb{P}(E_t)} = 1 - \frac{L_{\not{S}}(E_t)P_{\not{S}}(t)}{M(E_t)}$$

Method overview

1. $M(E_t)$ is the likelihood of observing E_t in general.
2. $L_{\not{S}}(E_t)$ is the likelihood of observing E_t provided $\not{S}$.
3. $P_{\not{S}}(t)$ is chosen to model the temporal seizure patterns.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. $L(E_t)$ and $M(E_t)$ are computed with Monte Carlo Quantile Estimation. The likelihood is the quantile of the sample, trimmed at the HDR$_\alpha$*boundary*.P(t)*isapriorfunctionwhichreliesonlyonthetimecovariate*.
2. ! Write Equivalent goal on board

Priors

Uniform prior

$$\mathbb{P}(S_t) = \lambda_0 \cdot T$$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. The uniform prior is the simplest. It means I give equal weight to each time point. My knowledge does not increase or degrade at any point.

Priors

Uniform prior

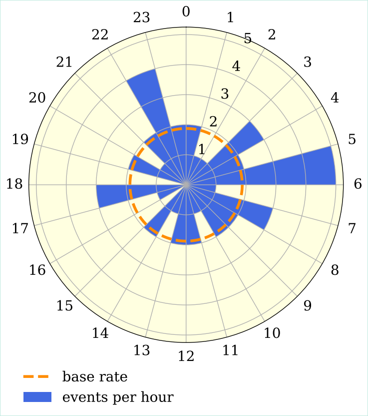
$$\mathbb{P}(S_t) = \lambda_0 \cdot T$$

Circadian prior

$$f(x|\mu, \kappa; \omega) = \frac{\exp(\kappa \cos(\omega(x - \mu)))}{2\pi I_0(\kappa)}$$

$$\mathbb{P}(S_t) = \frac{1}{K} \sum_{i=0}^{23} f(t \mid i, k)$$

Circadian prior



 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. The uniform prior is the simplest. It means I give equal weight to each time point. My knowledge does not increase or degrade at any point.
2. Another option is to look at seizure cycles, which has been explored more in recent years.
3. Following other works, we set the prior to be cyclical with the von Mises distribution. Where K is a normalizing constant evaluated numerically (`np.trapz`).
4. In this work, we set $\omega \leftarrow \frac{2\pi}{24}$ to scale the period to 24-hours, and drop it from the notation for brevity in the following text.

Priors

Uniform prior

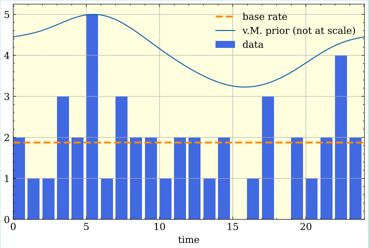
$$\mathbb{P}(S_t) = \lambda_0 \cdot T$$

Circadian prior

$$f(x|\mu,\kappa;\omega) = \frac{\exp(\kappa\cos(\omega(x-\mu)))}{2\pi I_0(\kappa)}$$

$$\mathbb{P}(S_t) = \frac{1}{K}\sum_{i=0}^{23} f(t \mid i, k)$$

Circadian prior



 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. The uniform prior is the simplest. It means I give equal weight to each time point. My knowledge does not increase or degrade at any point.
2. Another option is to look at seizure cycles, which has been explored more in recent years.
3. Following other works, we set the prior to be cyclical with the von Mises distribution. Where K is a normalizing constant evaluated numerically (`np.trapz`).
4. In this work, we set $\omega \leftarrow \frac{2\pi}{24}$ to scale the period to 24-hours, and drop it from the notation for brevity in the following text.

Learned embeddings

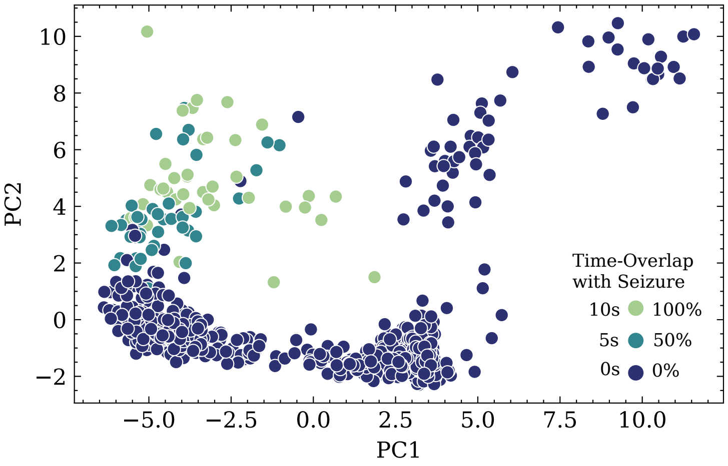


Figure 6: Embeddings portrayed in two dimensions.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. Each point represents a 10-second EEG segment fit with a Gaussian process.
2. The embeddings are learned via gradient descent on a variational lower bound. Maximizing this ensures that the Gaussian Process parameters learned are those that maximize the log-likelihood of the observed data.

Estimating densities

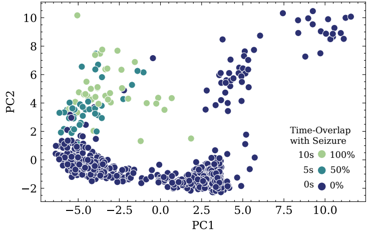


Figure 7: Embeddings portrayed in two dimensions.

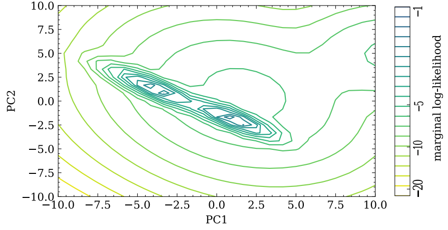
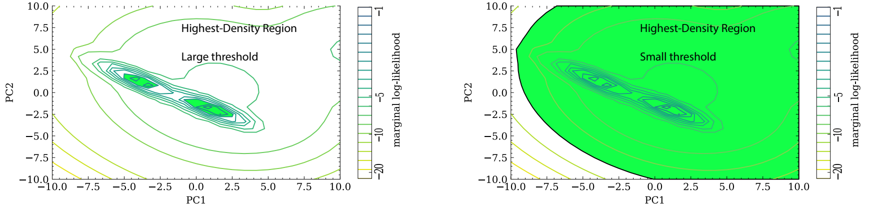


Figure 8: Fitting a GMM density estimator gives $\hat{pdf}(e)$

Side story: Highest density regions



Definition (α- Highest density region)

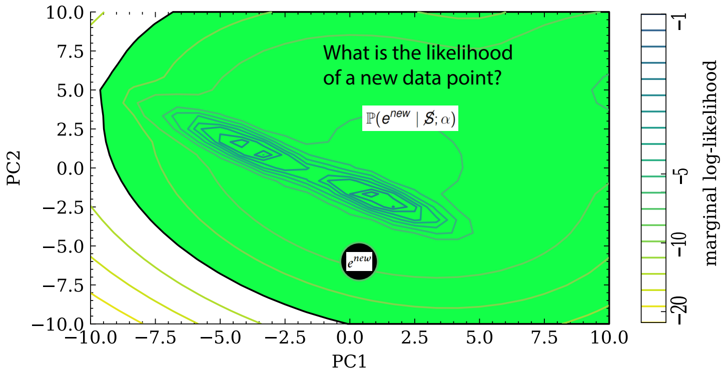
$$R(f_\alpha) = \{x : f(x) \geq f_\alpha\},$$

f_α is the largest constant such that $\mathbb{P}(X \in R(f_\alpha)) \geq 1 - \alpha$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

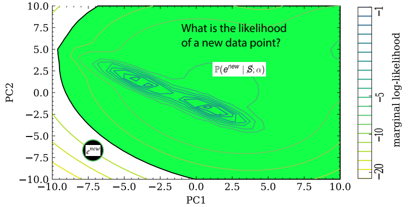
1. Highest density regions are regions in a space, such that a specific proportion of the mass lies within the region.
2. For example on the left, when we choose a low α, a low threshold for the highest density region, we get two small peaks which account for about 10 percent of the total mass.
3. On the example on the right, we choose a high α, so we get a large region which accounts for about 90 percent of the mass.
4. Now, this threshold α is a hyperparameter of the model. But let's assume it's fixed for the next slide, at a relatively low threshold, between 5 and 10 percent.

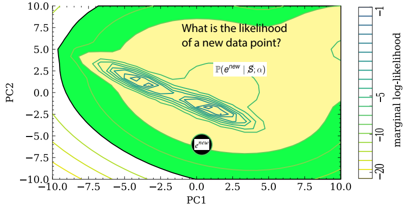
Likelihood function



1. The green is the 0.05 highest density region.

Likelihood function





Definition (BSLE likelihood)

$$L_{\mathcal{S}}(e^{new}) = \mathbb{P}(e^{new} \mid \mathcal{S}; \alpha) = \begin{cases} 0 & \text{if } e^{new} \notin HDR_\alpha \\ \mathcal{Z}(e^{new}) & \text{if } e^{new} \in HDR_\alpha \end{cases}$$

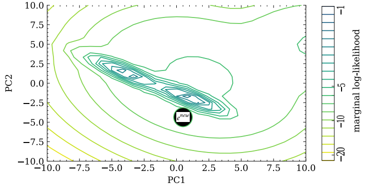
Marginal Likelihood

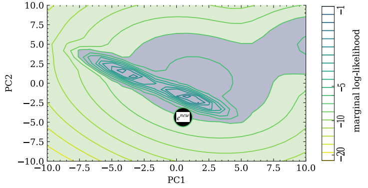
Definition (BSLE marginal likelihood)

$$\mathbb{P}(E) = \mathbb{P}(\{pdf(e) \leq pdf(E)\})$$

Geometric meaning

The probability of observing a sample e^{new}, is equal to the purple region divided by the entire region.



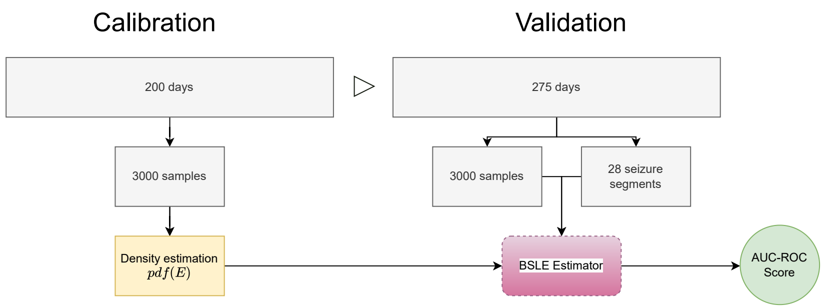


 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. The essence of the novelty score $\mathcal{Z}(e^{new})$ is that it maps data samples to the interval $[0, 1]$, such that more anomalous samples (i.e. samples from less dense regions) are given larger values.
2. We adopt quantiles as a way of measuring risk.
3. For a fixed constant $0 < p < 1$, the p-quantile of a continuous random variable is a constant ζ such that p of the distribution's mass lies below ζ.
4. For example, the median is the 0.5-quantile.
5. The method used here is called Quantile estimation with Monte Carlo Sampling.

1. Introduction
2. Research Problem
3. Research Plan
4. **Empirical Results**
5. Conclusion

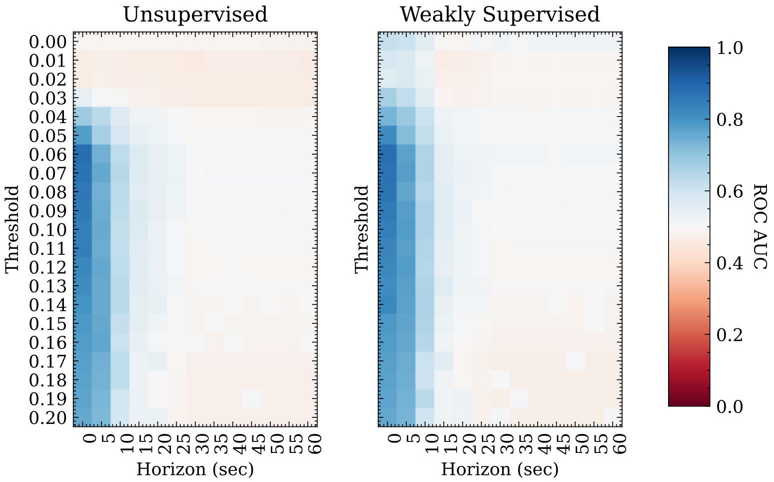
Simulating a prospective study



1. This separation of concerns ensures simulating a real-time setting, in which the 'offline' calibration phase is followed by an 'online' seizure detection mode.

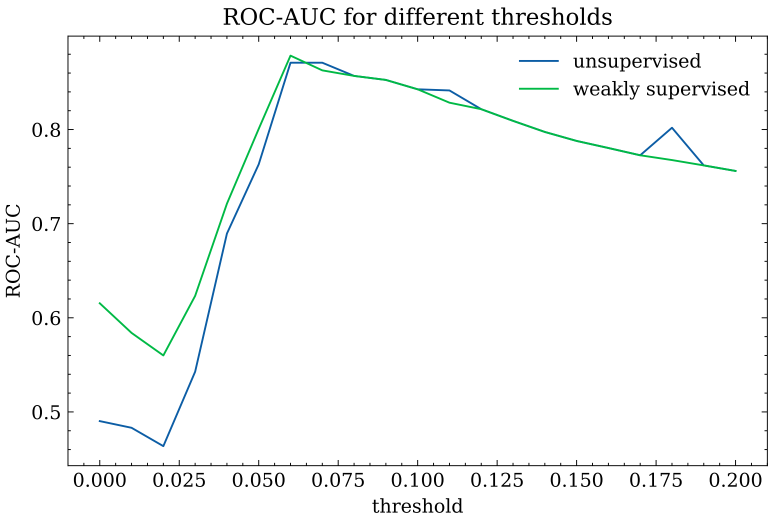
 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Evaluating BSLE on validation data



1. AUC-ROC scores smoothly distributed in the hyper-parameter space of both the unsupervised and weakly-supervised models, showing a low risk of over-fitting

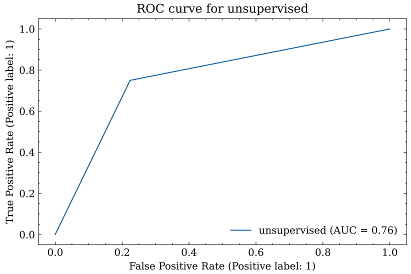
Evaluating BSLE on validation data

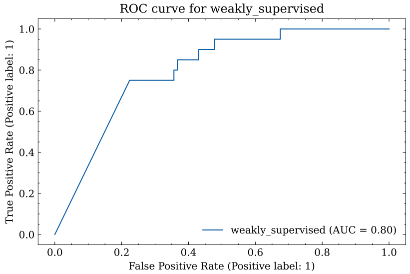


 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. AUC-ROC scores smoothly distributed in the hyper-parameter space of both the unsupervised and weakly-supervised models, showing a low risk of over-fitting

Evaluating BSLE on validation data





 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. AUC-ROC scores smoothly distributed in the hyper-parameter space of both the unsupervised and weakly-supervised models, showing a low risk of over-fitting

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a siezure is not occurring, and a temporal prior to induce bias based on labels, if available.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a siezure is not occurring, and a temporal prior to induce bias based on labels, if available.

BSLE can be used for unsupervised seizure alerts, and in addition learn from untaped seizures.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a siezure is not occurring, and a temporal prior to induce bias based on labels, if available.

BSLE can be used for unsupervised seizure alerts, and in addition learn from untaped seizures.

More work is required to statistically validate findings. Further work could use BSLE as a pseudo-labeler module in self-supervised learning.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

End

Thank You

References I

O. Deiss. *Efficient labeling technique and interpretable deep neural network for the classification of seizures using continuous electroencephalograms*. PhD thesis, Georgia Institute of Technology, 2018.

K. Saab, J. Dunnmon, C. Ré, D. Rubin, and C. Lee-Messer. Weak supervision as an efficient approach for automated seizure detection in electroencephalography. *NPJ digital medicine*, 3(1):1–12, 2020.

Y. Yang, N. D. Truong, J. K. Eshraghian, A. Nikpour, and O. Kavehei. Weak self-supervised learning for seizure forecasting: a feasibility study. *Royal Society Open Science*, 9(8):220374, 2022.

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

Apx. 1 - Gaussian processes

Definition (Gaussian process)

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

where

$$m(x) = \mathbb{E}[f(x)]$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License*.

1. A Gaussian process, defined by a mean function and covariance kernel function, with a Matérn kernel is a specific family of distribution over functions. Parameterized by <FILL IN>, the
2. d is the distance between x and x' scaled by the *lengthscale* parameter Φ
3. ν is a smoothness parameter. In this work, it is taken to be $\frac{2}{3}$, which limits us to functions which are exactly once differentiable.
4. K_ν is a modified Bessel function. The modified Bessel function has the property of being rapidly decaying, which means that it falls off quickly as the distance between two points increases.
5. +++This property is important in the context of kernel-based machine learning algorithms because it allows the kernel function to effectively capture the local structure of the data, while ignoring irrelevant or noisy features that may be present in the data.

Apx. 1 - Gaussian processes

Definition (Gaussian process)

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

where

$$m(x) = \mathbb{E}[f(x)]$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$$

Definition (Matérn class of covariance functions)

$$k_{Matern}(x, x') := \frac{2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}d)^{\nu} K_{\nu}(\sqrt{2\nu}d)$$

where

$$d = (x - x')^T \Phi^{-2} (x - x')$$

 Slides licensed under the *Creative Commons Attribution-ShareAlike 3.0 License.*

1. A Gaussian process, defined by a mean function and covariance kernel function, with a Matérn kernel is a specific family of distribution over functions. Parameterized by <FILL IN>, the
2. d is the distance between x and x' scaled by the *lengthscale* parameter Φ
3. ν is a smoothness parameter. In this work, it is taken to be $\frac{2}{3}$, which limits us to functions which are exactly once differentiable.
4. K_{ν} is a modified Bessel function. The modified Bessel function has the property of being rapidly decaying, which means that it falls off quickly as the distance between two points increases.
5. +++This property is important in the context of kernel-based machine learning algorithms because it allows the kernel function to effectively capture the local structure of the data, while ignoring irrelevant or noisy features that may be present in the data.