Towards interpretable automated seizure detection through Fast Gradient Boosting applied on EEG signals

Grascomp Doctoral Day 2019

Pierre De Handschutter

Joint work with P. Vanabelle, R. El Tahry, M. Boukhebouze and M. Benjelloun

Faculté Polytechnique
Université de Mons

Friday 22nd November 2019

What was the context of this work ?

What was the context of this work ?

- Internship in the research center CETIC

What was the context of this work ?

- Internship in the research center CETIC

- Collaboration with epileptologists from Saint-Luc Hospital

What was the context of this work ?

- Internship in the research center CETIC

Your Connection to ICT Research

- Collaboration with epileptologists from Saint-Luc Hospital

- Investigate "P4 medicine"

Figure taken from `https://bit.ly/2QpgraQ`

What was the context of this work ?

- Internship in the research center CETIC

- Collaboration with epileptologists from Saint-Luc Hospital

- Investigate "P4 medicine"

Figure taken from `https://bit.ly/2QpgraQ`

- Original paper : *(Vanabelle et al., 2019)*

What was the context of this work ?

- Internship in the research center CETIC

- Collaboration with epileptologists from Saint-Luc Hospital

- Investigate "P4 medicine"

Figure taken from `https://bit.ly/2QpgraQ`

- Original paper : *(Vanabelle et al., 2019)*
- Feel free to contact me at `pierre.dehandschutter@umons.ac.be`

Why is seizure detection so important ?

Why is seizure detection so important ?

- Epilepsy is a neurological disease affecting more than 50 millions of people around the world

Why is seizure detection so important ?

- Epilepsy is a neurological disease affecting more than 50 millions of people around the world
- Hard to detect, predict and cure ; causes not well understood

Why is seizure detection so important ?

- Epilepsy is a neurological disease affecting more than 50 millions of people around the world
- Hard to detect, predict and cure ; causes not well understood
- Current way to diagnose seizure : non-invasive EEG

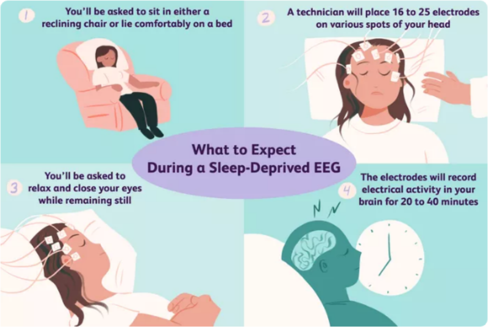


Figure taken from `https://bit.ly/2KtDugS`

Outline

Outline

Datasets

Datasets

- Finding good publicly available datasets is not a piece of cake !

Datasets

- Finding good publicly available datasets is not a piece of cake !
- Temple University Seizure Corpus (TUSZ) (v 1.2.1 ; April 18) : high diversity of patients, seizures, configurations, sampling rates... : 65GB, 266 patients in the training set (of whom 118 suffering from seizures), 50 in the test set (of whom 38 suffering from seizures), 40 different configurations of channels

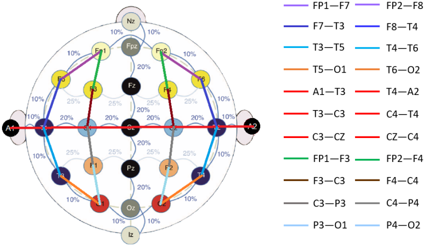
Datasets

- Finding good publicly available datasets is not a piece of cake !
- Temple University Seizure Corpus (TUSZ) (v 1.2.1 ; April 18) : high diversity of patients, seizures, configurations, sampling rates... : 65GB, 266 patients in the training set (of whom 118 suffering from seizures), 50 in the test set (of whom 38 suffering from seizures), 40 different configurations of channels
- A lot of models using Deep Learning techniques (CNN, LSTM,...) have been proposed but poor performance was reached *(Golmohammadi et al., 2017a)*, *(Golmohammadi et al., 2017b)*

What are the data ?

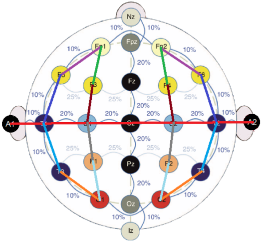
What are the data ?

- We use the 22 derivations, both transversal and longitudinal, from the Temporal Central Parasagittal (TCP) bipolar montage based on the 10/20 electrodes placement and either the Average Reference (AR) or Linked Ears (LE) reference



What are the data ?

- We use the 22 derivations, both transversal and longitudinal, from the Temporal Central Parasagittal (TCP) bipolar montage based on the 10/20 electrodes placement and either the Average Reference (AR) or Linked Ears (LE) reference

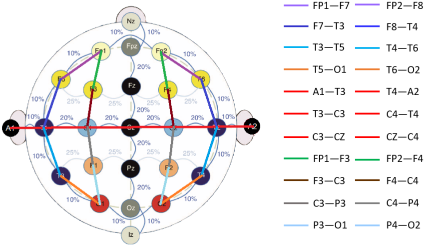


FP1—F7	FP2—F8
F7—T3	F8—T4
T3—T5	T4—T6
T5—O1	T6—O2
A1—T3	T4—A2
T3—C3	C4—T4
C3—CZ	CZ—C4
FP1—F3	FP2—F4
F3—C3	F4—C4
C3—P3	C4—P4
P3—O1	P4—O2

- Ten different types of seizures (7 generalized and 3 focal), focal seizures are major part of it

What are the data ?

- We use the 22 derivations, both transversal and longitudinal, from the Temporal Central Parasagittal (TCP) bipolar montage based on the 10/20 electrodes placement and either the Average Reference (AR) or Linked Ears (LE) reference



- Ten different types of seizures (7 generalized and 3 focal), focal seizures are major part of it
- Several kinds of labels annotations available ; we used aggregated binary annotations

Data preprocessing

- Computation of 22 features for each of the 22 channels, with PyEEG library, based on literature review

Data preprocessing

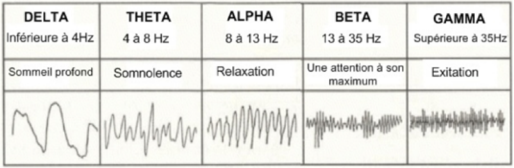
- Computation of 22 features for each of the 22 channels, with PyEEG library, based on literature review
  - Statistical features : moments of order up to 4, max, min,...

Data preprocessing

- Computation of 22 features for each of the 22 channels, with PyEEG library, based on literature review
  - Statistical features : moments of order up to 4, max, min,...
  - Other temporal features : Hjorth complexity and mobility, Petrosian fractal

Data preprocessing

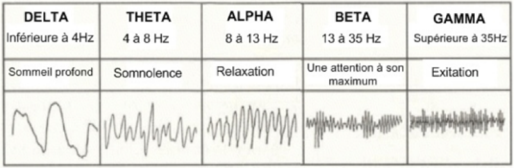
- Computation of 22 features for each of the 22 channels, with PyEEG library, based on literature review
  - Statistical features : moments of order up to 4, max, min,...
  - Other temporal features : Hjorth complexity and mobility, Petrosian fractal
  - Frequential features : power spectral density in the five frequency rhythms (alpha $(8-13\text{Hz})$, beta $(13-35\text{Hz})$, gamma $(>35\text{Hz})$, delta $(1-4\text{Hz})$, theta $(4-8\text{Hz})$) and the corresponding ratios, spectral centroid and monotony.



DELTA Inférieure à 4Hz	THETA 4 à 8 Hz	ALPHA 8 à 13 Hz	BETA 13 à 35 Hz	GAMMA Supérieure à 35Hz
Sommeil profond	Somnolence	Relaxation	Une attention à son maximum	Exitation

Data preprocessing

- Computation of 22 features for each of the 22 channels, with PyEEG library, based on literature review
  - Statistical features : moments of order up to 4, max, min,...
  - Other temporal features : Hjorth complexity and mobility, Petrosian fractal
  - Frequential features : power spectral density in the five frequency rhythms (alpha ($8 - 13$Hz), beta ($13 - 35$Hz), gamma (> 35Hz), delta ($1 - 4$Hz), theta ($4 - 8$Hz)) and the corresponding ratios, spectral centroid and monotony.



- These features are computed on non-overlapping segments of $1s$

Gradient Boosting in a nutshell

Gradient Boosting in a nutshell

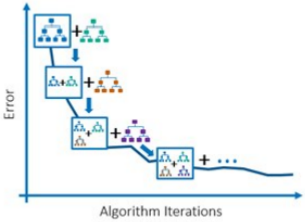
- Based on boosting i.e. sequential iterative models

Gradient Boosting in a nutshell

- Based on boosting i.e. sequential iterative models
- Start by fitting a classification decision tree to the original data and compute the misclassified points (residuals)

Gradient Boosting in a nutshell

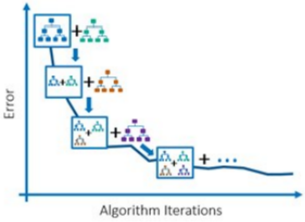
- Based on boosting i.e. sequential iterative models
- Start by fitting a classification decision tree to the original data and compute the misclassified points (residuals)
- At each iteration, fit a decision tree to the residuals and "add" it to the current model



https ://bit.ly/32W7Mzr

Gradient Boosting in a nutshell

- Based on boosting i.e. sequential iterative models
- Start by fitting a classification decision tree to the original data and compute the misclassified points (residuals)
- At each iteration, fit a decision tree to the residuals and "add" it to the current model



https ://bit.ly/32W7Mzr

- In our model, the maximal depth of a tree is set to 3 and the maximum number of iterations (i.e. of built trees) to 400

Particularities of XGBoost

1. Both figures taken from https://bit.ly/2OjULdR

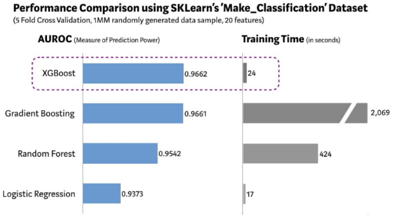
Particularities of XGBoost

- Scalable, Portable and Distributed Gradient Boosting

1. Both figures taken from `https://bit.ly/2OjULdR`

Particularities of XGBoost

- Scalable, Portable and Distributed Gradient Boosting
- Several hardware and algorithmic improvements compared to classical Gradient Boosting :

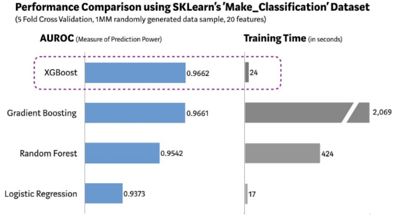


[1]

1. Both figures taken from https://bit.ly/2OjULdR

Particularities of XGBoost

- Scalable, Portable and Distributed Gradient Boosting
- Several hardware and algorithmic improvements compared to classical Gradient Boosting :



[1]

- Tree training and prediction can be accelerated with CUDA-capable GPUs (ex : *gpu_hist*).

1. Both figures taken from `https://bit.ly/2OjULdR`

First method : standard partitioning

First method : standard partitioning

- Split the train set into 80% of training data and 20% of validation

First method : standard partitioning

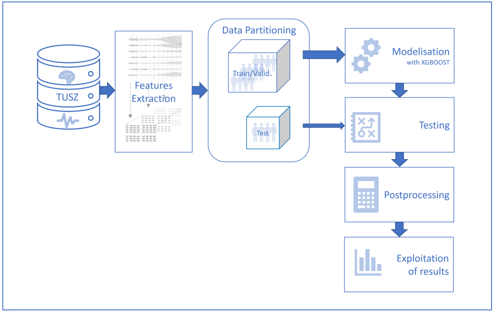
- Split the train set into 80% of training data and 20% of validation
- Train a XGBoost classifier of depth 3 with 400 estimators (taking into account class imbalance !)

First method : standard partitioning

- Split the train set into 80% of training data and 20% of validation
- Train a XGBoost classifier of depth 3 with 400 estimators (taking into account class imbalance !)
- Predict the label (seizure or background) of each segment of the test set

First method : standard partitioning

- Split the train set into 80% of training data and 20% of validation
- Train a XGBoost classifier of depth 3 with 400 estimators (taking into account class imbalance !)
- Predict the label (seizure or background) of each segment of the test set
- Compute performance metrics



Second method : Leave One Out Cross Test (LOOCT)

Second method : Leave One Out Cross Test (LOOCT)

- Choose a patient from the test set and "leave it out"

Second method : Leave One Out Cross Test (LOOCT)

- Choose a patient from the test set and "leave it out"
- Split all the other recordings (train and test set) between training and validation (80-20)

Second method : Leave One Out Cross Test (LOOCT)

- Choose a patient from the test set and "leave it out"
- Split all the other recordings (train and test set) between training and validation (80-20)
- Train a XGBoost classifier of depth 3 with 400 estimators

Second method : Leave One Out Cross Test (LOOCT)

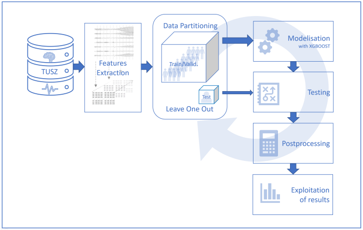
- Choose a patient from the test set and "leave it out"
- Split all the other recordings (train and test set) between training and validation (80-20)
- Train a XGBoost classifier of depth 3 with 400 estimators
- Predict the label (seizure or background) of each segment of the patient left out

Second method : Leave One Out Cross Test (LOOCT)

- Choose a patient from the test set and "leave it out"
- Split all the other recordings (train and test set) between training and validation (80-20)
- Train a XGBoost classifier of depth 3 with 400 estimators
- Predict the label (seizure or background) of each segment of the patient left out
- Repeat for each patient in the test set

Second method : Leave One Out Cross Test (LOOCT)

- Choose a patient from the test set and "leave it out"
- Split all the other recordings (train and test set) between training and validation (80-20)
- Train a XGBoost classifier of depth 3 with 400 estimators
- Predict the label (seizure or background) of each segment of the patient left out
- Repeat for each patient in the test set
- Compute performance metrics (average over all the patients in the test set)



Metrics

Metrics

- Confusion matrix for seizure detection :

		Predicted class	
		Background	Seizure
Actual class	Background	True Negatives (TN)	False Positives (FP)
	Seizure	False Negatives (FN)	True positives (TP)

Metrics

- Confusion matrix for seizure detection :

		Predicted class	
		Background	Seizure
Actual class	Background	True Negatives (TN)	False Positives (FP)
	Seizure	False Negatives (FN)	True positives (TP)

- Important metrics :

Metrics

- Confusion matrix for seizure detection :

Predicted class

Actual class		Background	Seizure
	Background	True Negatives (TN)	False Positives (FP)
	Seizure	False Negatives (FN)	True positives (TP)

- Important metrics :
  - Specificity $=\frac{TN}{TN+FP}$=1-false alarm rate

Metrics

- Confusion matrix for seizure detection :

		Predicted class	
		Background	Seizure
Actual class	Background	True Negatives (TN)	False Positives (FP)
	Seizure	False Negatives (FN)	True positives (TP)

- Important metrics :
  - Specificity $=\frac{TN}{TN+FP}$=1-false alarm rate
  - Sensitivity$=\frac{TP}{TP+FN}$

Metrics

- Confusion matrix for seizure detection :

Predicted class

Actual class	Background	Seizure
Background	True Negatives (TN)	False Positives (FP)
Seizure	False Negatives (FN)	True positives (TP)

- Important metrics :
  - Specificity $=\frac{TN}{TN+FP}$=1-false alarm rate
  - Sensitivity$=\frac{TP}{TP+FN}$
- Note that in the literature, the metrics were computed permissively "event-wise", this is the Any-Overlap method *(Ziyabari et al., 2017)*

Metrics

- Confusion matrix for seizure detection :

Predicted class

Actual class	Background	Seizure
Background	True Negatives (TN)	False Positives (FP)
Seizure	False Negatives (FN)	True positives (TP)

- Important metrics :
  - Specificity $=\frac{TN}{TN+FP}$=1-false alarm rate
  - Sensitivity$=\frac{TP}{TP+FN}$
- Note that in the literature, the metrics were computed permissively "event-wise", this is the Any-Overlap method *(Ziyabari et al., 2017)*
- Even with those metrics, specificity is generally not higher than 70/80% while sensitivity is always lower than 40%.

Outline

Experiments and results

Experiments and results

- With the AR montage only, LOOCT allows to increase the sensitivity (35.8% $\rightarrow$ 52.66%) but reduces the specificity (92.46% $\rightarrow$ 86.50%) compared to a standard partitioning

Experiments and results

- With the AR montage only, LOOCT allows to increase the sensitivity (35.8% $\rightarrow$ 52.66%) but reduces the specificity (92.46% $\rightarrow$ 86.50%) compared to a standard partitioning
- If we add the LE montages in the training set, the performance increases for the standard partitioning but few improvement is observed for LOOCT :

	AR Train	**LE Train**	**AR Test**	**1st meth.**	**2nd meth.**	**Sens.**	**Spec.**
1	train		test	x		35,80%	92,46%
2	train	train	test	x		41,74%	92,50%
3	train		train-test		x	52,66%	86,50%
4	train	train	train-test		x	53,87%	86,20%

Experiments and results

- With the AR montage only, LOOCT allows to increase the sensitivity (35.8% $\rightarrow$ 52.66%) but reduces the specificity (92.46% $\rightarrow$ 86.50%) compared to a standard partitioning
- If we add the LE montages in the training set, the performance increases for the standard partitioning but few improvement is observed for LOOCT :

	AR Train	LE Train	AR Test	1st meth.	2nd meth.	Sens.	Spec.
1	train		test	x		35,80%	92,46%
2	train	train	test	x		41,74%	92,50%
3	train		train-test		x	52,66%	86,50%
4	train	train	train-test		x	53,87%	86,20%

- Experiments on the LE_test :

Experiments and results

- With the AR montage only, LOOCT allows to increase the sensitivity (35.8% $\rightarrow$ 52.66%) but reduces the specificity (92.46% $\rightarrow$ 86.50%) compared to a standard partitioning
- If we add the LE montages in the training set, the performance increases for the standard partitioning but few improvement is observed for LOOCT :

	AR Train	**LE Train**	**AR Test**	**1st meth.**	**2nd meth.**	**Sens.**	**Spec.**
1	train		test	x		35,80%	92,46%
2	train	train	test	x		41,74%	92,50%
3	train		train-test		x	52,66%	86,50%
4	train	train	train-test		x	53,87%	86,20%

- Experiments on the LE_test :
  - Harder to interpret (the sensitivities are way higher than with the AR montages but the specificities a bit lower)

Experiments and results

- With the AR montage only, LOOCT allows to increase the sensitivity (35.8% $\rightarrow$ 52.66%) but reduces the specificity (92.46% $\rightarrow$ 86.50%) compared to a standard partitioning
- If we add the LE montages in the training set, the performance increases for the standard partitioning but few improvement is observed for LOOCT :

	AR Train	LE Train	AR Test	1st meth.	2nd meth.	Sens.	Spec.
1	train		test	x		35,80%	92,46%
2	train	train	test	x		41,74%	92,50%
3	train		train-test		x	52,66%	86,50%
4	train	train	train-test		x	53,87%	86,20%

- Experiments on the LE_test :
  - Harder to interpret (the sensitivities are way higher than with the AR montages but the specificities a bit lower)
  - The low SNR of AR signals might explain the poor results obtained when AR recordings are used in the training set

Analysis by type of seizures

Analysis by type of seizures

- Sensitivities :

	AR Train	LE Train	AR Test	1st meth.	2nd meth.	Sens. foc.	Sens. gen
1	train		test	x		23,02%	59,50%
2	train	train	test	x		27,68%	66,24%
3	train		train-test		x	35,02%	81,51%
4	train	train	train-test		x	37,61%	81,10%

Analysis by type of seizures

- Sensitivities :

	AR Train	LE Train	AR Test	1st meth.	2nd meth.	Sens. foc.	Sens. gen
1	train		test	x		23,02%	59,50%
2	train	train	test	x		27,68%	66,24%
3	train		train-test		x	35,02%	81,51%
4	train	train	train-test		x	37,61%	81,10%

- Generalized seizures way better detected than focal ones, as the seizure information is widespread over all the channels

Analysis by type of seizures

- Sensitivities :

	AR Train	LE Train	AR Test	1st meth.	2nd meth.	Sens. foc.	Sens. gen
1	train		test	x		23,02%	59,50%
2	train	train	test	x		27,68%	66,24%
3	train		train-test		x	35,02%	81,51%
4	train	train	train-test		x	37,61%	81,10%

- Generalized seizures way better detected than focal ones, as the seizure information is widespread over all the channels

- This would advocate for the design of type-specific models but the types of seizures are not independent (several types of seizures can appear in the same recording, focal seizures might "generalize",...)

Analysis of features importance

Analysis of features importance

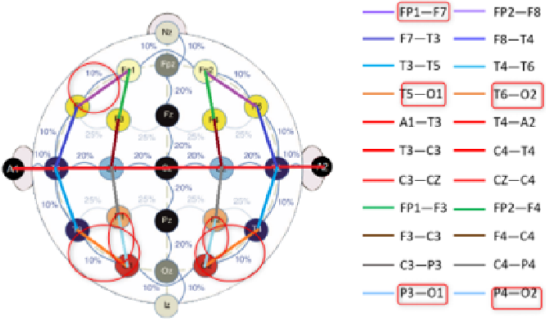
- XGBoost is a tree-based decision model so it allows to retrieve the features importance

Analysis of features importance

- XGBoost is a tree-based decision model so it allows to retrieve the features importance
- Interpretability is critical in the medical domain *(Itani et al., 2018)*

Analysis of features importance

- XGBoost is a tree-based decision model so it allows to retrieve the features importance
- Interpretability is critical in the medical domain *(Itani et al., 2018)*
- Power Spectral Density in the α, β, θ bands of some channels of the posterior region of the brain seem the most discriminative



Analysis of features importance (2)

Two possible explanations from a medical point-of-view :

Analysis of features importance (2)

Two possible explanations from a medical point-of-view :

- Focal seizures of occipital origin (although there are rare) start there and some generalized seizures induce changes in this part as well

Analysis of features importance (2)

Two possible explanations from a medical point-of-view :

- Focal seizures of occipital origin (although there are rare) start there and some generalized seizures induce changes in this part as well
- The occipital region is the one where a visual discrimination of ictal and interictal phases is the easiest to (empirically) detect, according to neurologists $\Rightarrow$ the algorithm maybe found the same rule !

Analysis of features importance (2)

Two possible explanations from a medical point-of-view :

- Focal seizures of occipital origin (although there are rare) start there and some generalized seizures induce changes in this part as well
- The occipital region is the one where a visual discrimination of ictal and interictal phases is the easiest to (empirically) detect, according to neurologists $\Rightarrow$ the algorithm maybe found the same rule !
- The second hypothesis seems confirmed by some other recent studies *(Shah et al., 2017)* but further research should validate this intuition

Outline

Conclusion and perspectives

- Conclusion :

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative
- Perspectives :

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative
- Perspectives :
  - Test our model on more recent released (cleaner, broader) versions of the TUSZ

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative
- Perspectives :
  - Test our model on more recent released (cleaner, broader) versions of the TUSZ
  - Develop a multi-class classification model

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative
- Perspectives :
  - Test our model on more recent released (cleaner, broader) versions of the TUSZ
  - Develop a multi-class classification model
  - Focus on detecting the location and time of seizures onsets

Conclusion and perspectives

- Conclusion :
  - Data quality and quantity are critical in seizure detection problems
  - A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
  - The increase of training data allows higher performance but the partitioning of data is key to get meaningful results
  - Some channels and features seem more discriminative
- Perspectives :
  - Test our model on more recent released (cleaner, broader) versions of the TUSZ
  - Develop a multi-class classification model
  - Focus on detecting the location and time of seizures onsets
  - Use other sources of signal (ECG,...)

References I

Meysam Golmohammadi, Saeedeh Ziyabari, Vinit Shah, Silvia Lopez de Diego, Iyad Obeid, and Joseph Picone. Deep architectures for automated seizure detection in scalp eegs. *arXiv preprint arXiv :1712.09776*, 2017a.

Meysam Golmohammadi, Saeedeh Ziyabari, Vinit Shah, Eva Von Weltin, Christopher Campbell, Iyad Obeid, and Joseph Picone. Gated recurrent networks for seizure detection. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2017 IEEE*, pages 1–5. IEEE, 2017b.

Sarah Itani, Fabian Lecron, and Philippe Fortemps. Specifics of medical data mining for diagnosis aid : A survey. *Expert Systems with Applications*, 2018.

References II

Vinit Shah, Meysam Golmohammadi, Saeedeh Ziyabari, Eva Von Weltin, Iyad Obeid, and Joseph Picone. Optimizing channel selection for seizure detection. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5. IEEE, 2017.

Paul Vanabelle, Pierre De Handschutter, Riëm El Tahry, Mohammed Benjelloun, and Mohamed Boukhebouze. Epileptic seizure detection using eeg signals and extreme gradient boosting. *Journal of Biomedical Research*, 2019.

Saeedeh Ziyabari, Vinit Shah, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. Objective evaluation metrics for automatic classification of eeg events. *arXiv preprint arXiv :1712.10107*, 2017.

Towards interpretable automated seizure detection through Fast Gradient Boosting applied on EEG signals

Grascomp Doctoral Day 2019

Pierre De Handschutter

Joint work with P. Vanabelle, R. El Tahry, M. Boukhebouze and M. Benjelloun

Faculté Polytechnique
Université de Mons

Friday 22nd November 2019

Datasets

Datasets

- Finding good publicly available datasets is not a piece of cake !

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation
  - CHB-MIT dataset : only children, annotations specifically designed for prediction tasks

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation
  - CHB-MIT dataset : only children, annotations specifically designed for prediction tasks
  - Temple University Seizure Corpus (TUSZ) : high diversity of patients, seizures, configurations, sampling rates... $\Rightarrow$ the one we used

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation
  - CHB-MIT dataset : only children, annotations specifically designed for prediction tasks
  - Temple University Seizure Corpus (TUSZ) : high diversity of patients, seizures, configurations, sampling rates... $\Rightarrow$ the one we used
- TUSZ dataset v 1.2.1 (April 18) :

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation
  - CHB-MIT dataset : only children, annotations specifically designed for prediction tasks
  - Temple University Seizure Corpus (TUSZ) : high diversity of patients, seizures, configurations, sampling rates... $\Rightarrow$ the one we used
- TUSZ dataset v 1.2.1 (April 18) :
  - Some statistics : 65GB, 266 patients in the training set (of whom 118 suffering from seizures), 50 in the test set (of whom 38 suffering from seizures), 40 different configurations of channels

Datasets

- Finding good publicly available datasets is not a piece of cake !
  - Bonn dataset : short, single-channel EEG without appropriate documentation
  - CHB-MIT dataset : only children, annotations specifically designed for prediction tasks
  - Temple University Seizure Corpus (TUSZ) : high diversity of patients, seizures, configurations, sampling rates... ⇒ the one we used
- TUSZ dataset v 1.2.1 (April 18) :
  - Some statistics : 65GB, 266 patients in the training set (of whom 118 suffering from seizures), 50 in the test set (of whom 38 suffering from seizures), 40 different configurations of channels
  - A lot of models using Deep Learning techniques (CNN, LSTM,...) have been proposed but poor performance was reached *(Golmohammadi et al., 2017a)*, *(Golmohammadi et al., 2017b)*

What are the data ?

What are the data ?

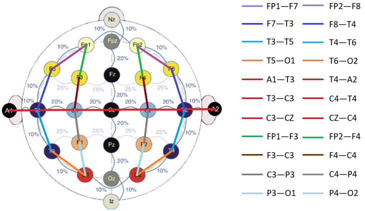
- Despite the variety of channels configurations, all the recordings contain the electrodes of the standard 10/20 placement

What are the data ?

- Despite the variety of channels configurations, all the recordings contain the electrodes of the standard 10/20 placement
- EDF files contain what is recorded in practice i.e. the difference between the signal recorded at each electrode of the 10/20 placement and a reference (unipolar montage). Two references are considered in TUSZ : either Average Reference (AR) or Linked Ears (LE)

What are the data ?

- Despite the variety of channels configurations, all the recordings contain the electrodes of the standard 10/20 placement
- EDF files contain what is recorded in practice i.e. the difference between the signal recorded at each electrode of the 10/20 placement and a reference (unipolar montage). Two references are considered in TUSZ : either Average Reference (AR) or Linked Ears (LE)
- Neurologists prefer to use bipolar montages, to take into account neighbourhood information. The Temporal Central Parasagittal (TCP) bipolar montage is used and contains 22 derivations, both transversal and longitudinal



More about the data...

More about the data...

- There are 10 different kinds of seizures inside the dataset (7 of generalized seizures and 3 of focal seizures), focal seizures are major part of it

More about the data...

- There are 10 different kinds of seizures inside the dataset (7 of generalized seizures and 3 of focal seizures), focal seizures are major part of it
- There are four kinds of labels annotations available :

More about the data...

- There are 10 different kinds of seizures inside the dataset (7 of generalized seizures and 3 of focal seizures), focal seizures are major part of it
- There are four kinds of labels annotations available :
  - Channel-by-channel annotations indicating the start and end of each seizure (by type) in the recording in each channel and its binary version (seizure vs background)

More about the data...

- There are 10 different kinds of seizures inside the dataset (7 of generalized seizures and 3 of focal seizures), focal seizures are major part of it
- There are four kinds of labels annotations available :
  - Channel-by-channel annotations indicating the start and end of each seizure (by type) in the recording in each channel and its binary version (seizure vs background)
  - Event-based annotations indicating the start and end of each seizure (by type) in the recording overall and its binary version (seizure vs background)

More about the data...

- There are 10 different kinds of seizures inside the dataset (7 of generalized seizures and 3 of focal seizures), focal seizures are major part of it
- There are four kinds of labels annotations available :
  - Channel-by-channel annotations indicating the start and end of each seizure (by type) in the recording in each channel and its binary version (seizure vs background)
  - Event-based annotations indicating the start and end of each seizure (by type) in the recording overall and its binary version (seizure vs background)
- The binary event-based annotations are the most commonly used in ML.

Motivation of the proposed approach

Motivation of the proposed approach

- Work directly on raw data is not easy since the amount of data is too high and there are artefacts $\Rightarrow$ features-based approach

Motivation of the proposed approach

- Work directly on raw data is not easy since the amount of data is too high and there are artefacts $\Rightarrow$ features-based approach
- A generic modeling seems more useful than a patient-specific approach

Motivation of the proposed approach

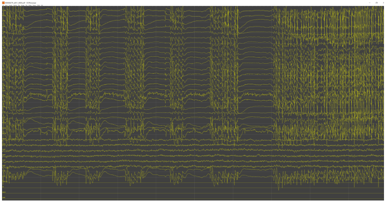
- Work directly on raw data is not easy since the amount of data is too high and there are artefacts $\Rightarrow$ features-based approach
- A generic modeling seems more useful than a patient-specific approach
- Deep learning models are hard to design, little interpretable and not convincing in the recent literature

Motivation of the proposed approach

- Work directly on raw data is not easy since the amount of data is too high and there are artefacts $\Rightarrow$ features-based approach
- A generic modeling seems more useful than a patient-specific approach
- Deep learning models are hard to design, little interpretable and not convincing in the recent literature
- XGBoost is a fast implementation of the Gradient Boosting algorithm, which is interpretable

Experiments and results (2)

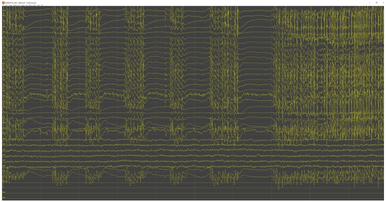
- The precision$=\frac{TP}{TP+FP}$ is quite low $(30-40\%)$ and might be partially due to ambiguous labelling



Repetitive inter-ictal generalized spike and wave discharges alternated with global EEG signal attenuation, followed by a generalized seizure

Experiments and results (2)

- The precision$=\frac{TP}{TP+FP}$ is quite low $(30-40\%)$ and might be partially due to ambiguous labelling

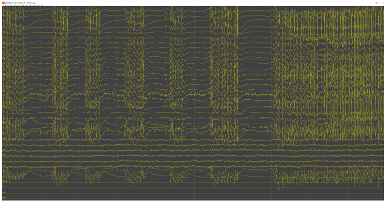


Repetitive inter-ictal generalized spike and wave discharges alternated with global EEG signal attenuation, followed by a generalized seizure

- A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered

Experiments and results (2)

- The precision$=\frac{TP}{TP+FP}$ is quite low $(30-40\%)$ and might be partially due to ambiguous labelling



Repetitive inter-ictal generalized spike and wave discharges alternated with global EEG signal attenuation, followed by a generalized seizure

- A global model might be too ambitious and does not take into the account the particularities of the types of seizures encountered
- In summary, the increase of training data allows higher performance but the partitioning of data is key to get meaningful results