Classification of epilepsy ictals with information content based on contrast model using GAN

Qin He 1st Affiliation (Author)
Data Engineering and Machine Learning
Signal Processing
Tampere University of Science and Technology
Tampere, Finland
qin.he@tuni.fi

Jari Hyttinen 2nd Affiliation (Author)
Biophysics
Tampere University of Science and Technology
Tampere, Finland
jari.hyttinen@tuni.fi

Sampa Pursiainen 2nd Affiliation (Author)
Applied Mathematics
Tampere University of Science and Technology
Tampere, Finland
Sampsa.pursiainen@tuni.fi

Jukka Peltola 2nd Affiliation (Author)
Health and Life Science
Tampere University and University Hospital
Tampere, Finland
Jukka.Peltola@tuni.fi

Xiaochuan Pan 2nd Affiliation (Author)
Maths and Applied Maths
East China University and Science and Technology
Shanghai, China
pxc@ecust.edu.cn

RubinWang 2nd Affiliation (Author)
Maths and Applied Maths
East China University and Science and Technology
Shanghai, China
RubinWang@ecust.edu.cn

***Abstract*— Ontology mining is one of the most useful mechanisms. It is most commonly used on the application constructed on database and in vision of the semantic (Web) distributed knowledge bases. As one of the character of the Web data is the possibility to have new resources in a context of the open-world semantics, we thus apply it on the ictal data, which also break out spontaneously from dynamically changing sources. When dealing with this kind of data, the GAN is quite useful as it generate adversary data automatically to update computing on detecting features with better performance. In addition, as the classification is conducted on its structural information, the ontology is useful in the case when the important properties of the ictals are listed. Under such data, our classification is possible to achieve good accuracy targeting their similarity. Contrast models are commonly used and in our case, either continuous features or discrete knowledge base can be utilized.**

Keywords-Data mining and Machine Learning, Knowledge Base, Ictal Classification, GAN, Competative Reward Evaluation

I. INTRODUCTION

Information theory is commonly used in communication, data mining, machine learning and many interdisciplinary researches, combining modeling and computation from basic operations with logic and functions to complex problems on higher ordered systems. As Shannon Claude constructed the theory to communication operations and signal process[1], a variety of related models are built on different fields, including the linguistics study, universe big bang theory, numerous approximation and simulations and etc. In this paper, the idea origin from semantics web (SW) is further developed with the features extracted after intrinsic mode function (IMF) decomposed the ictal signals, and its advantage of dealing with instant updated information is also utilized with the deep learning model generative adversarial network (GAN)[2,3], classifying the ictal signal of multi sources, inferencing the type of ictal as a binary classification problem.

Specifically, in the feature extraction, the ictal time series is transformed onto frequency domain for retrieving related entropy. On one hand, the features are from both data generated through generator and evaluated through discriminator, and the type of the features are originally on continuous, real domain. To deal with the critical components of the signal, angle, phase and power are all computed stored with its median, minimum and maximum. To conduct further analysis, first order logic(FOL) is applied quantifying the relation of the property. For interictal, it is stored as 1 if the range of the signal is larger than the adversary signal, the same rule is constructed for phase and power. However, for the median is also of interest, it is stored as 1 if the value of phase is larger than the adversary compartment while for the median of power, it reversely is stored as 1 if it is smaller. Vice versa for the ictal signals. Such fragment of FOL is well-known as the intersection of inductive logic programming(ILP)[4], data mining and knowledge reasoning and representation[5]. As the model adopted is of higher order on basis of distance, the Hamiltonians computed and used directly in the coarse network is trained with logistic regression. Note that, the information content(IC) are given by logP and thus for convenience, we have the features normalized into (0,1] first. Naturally, it gives out the map onto log distributed scale. As mentioned, the position quantities and velocity quantities are computed through the entropy combining instantaneous energy and instantaneous frequency as well as the difference combing (IMF)[6] as well which also requires the decomposition utilizing envelope, named empirical mode decomposition(EMD)[7], which is commonly used in signal filter as a heuristic model.

On the other hand, the distance computation in the experiment sums up the ICs linearly. Intuitively, as the ICs are log value, analog to the origin theory which exclude the computation on expressive language(dealing with is a problem.) this model is not suitable for features in binary coded or with zero crossed values. Under such circumstances, we need to do nan replace process as the EEG signal do contain quite many zero values.

As mentioned, the dynamical process is of interest and in addition to the instantaneous quantities in the pre-computed net, the feature matching based on contrast model is updated dynamically on each step first evaluated through discriminator and then the generator. However, to avoid the large computation, the variation inference[8] is conducted to regenerate the adversary signal and put inn use for training instead of being regenerated in every epoch. And consider the character of the ictal signal, the generation combines the gamma distribution, which is in consistent with the poisson process and the time point is also restored in the table. To guarantee the convergence of the system, the reward function is set ass a zero-sum function with 1 point for ictal related feature and -1 for interictal related feature.

Meanwhile, the experiment part test the data with distance relation and thus applies the linear discriminate model(LDA) in optimizing the likelihood and the evaluation for prediction is the competitive reward evaluation function[9].

In the rest of the paper, the structures are arranged as following, the structure of the data and some importantmodel parameters are described in the chapter 2, the feature matching contrast model using GAN is proposed in chapter 3 along with its algorithm application, the chapter 4 contains the experiment and result analysis and the further discussion and future direction is in the last part chapter 5. Tables are included in with figures and demonstration in Appendix along with related previous work at[16]

II. DASTUCTURE AND MODEL PARAMETERS

Data used in this paper is the 64 channel ictal EEG data(120000 samples per channel achieved with sampling frequency 2000 Hz.).(Detail of 5 ictal, variable: ictal_n_interv, which is segmented according to the manually annotated labels in the previous work sees appendix A.)

Features are in 'Example.mat' can be downloaded from:

The Example structures contains angle, phase, power, Hamiltonians of original white noise added 5th ictal,and the generated adversary compartment in X1,labeled {-1,1} in Y1 another track X2 sampled with same condition and environment but annotated as {1, -1} in Y2. Both the tracks are made as table format.

Parameters from Hamiltonians pre-computed network through Discriminator are stored in X1.(10:18,1) while those through Generator are stored in X1(10:18,2).Some temporary but might be useful quantities after EMD about Similarly, IMFs are stored in row numbers 1:9 for both D and G blocks. Training is with window size 128 in pre-processing, bin size 2000 samples/ms and batch size [1,2, 5,10,100] with steps [10,50]. (Example shows 50 epochs in each 5 sized batch.

III. ASSEMBEL MODEL PROPOSED AND ALGORITHM APPLICATION

The Hamiltonians of position and velocity information computation is mostly introduced in the previous work[11], in this paper, the temp process is simply iteration with SGD. In the leapfrog process, log ratio, for instance, introduced by Dino Sejdinovic[10]. As the ictal signal is usually of high noise, each epoch in the discriminator before giving reward and discriminate the signal type, there will firstly add white noise to the signal and which is believed to cancel the Bayes inference through randomization.

Evaluation with reward:

And in our paper, the reward function is Z based instead of W based::

with the distribution f: lpdf, R average value of: $lambda = m +_2^{\eta} \frac{g\,lpdf}{lpdf}$

and any step k, signal x and exclusive $x' = X\backslash\{x\}$, we have:

1. non-learnable level: $\frac{\partial Rx^f}{\partial \prod x} = 0$
2. Isolated level : $\frac{\partial Rx^f}{\partial \prod x'} = 0$
3. Competitive level: $\frac{\partial \int_{x \neq x} Rx'^f dx'}{\partial \prod x'} = 0$
4. Collaborative level: $\frac{\partial Rx'^f}{\partial Rx^f} \geq 0$
5. Competitive and Collaborative Mixed: $\int_{x \neq x} \partial Rx'^f dx' / \partial \prod x' = 0$

A. *Pre-network in Discriminator*

First, mix signal with white noise and compute the Hamiltonians for further classification.

G<-output z,ld -(when s=0)-dx/x<min(exp(lambda),1)?|replace p <-config white noise+ | SE->p | |
Input(EEG_D0)-→IMFs-→ instantanious f, Ek->U-> q->z->lambda->|

Algorithm I(Block D)
While s >0 , do:
1. signal = signal + white noise
2. decompose signal into IMFs(detail see previous work[11]), formula
C. Algorithm: likelihood-free HMCS on IMF/EMD step 1-12

Note that: for single competitive case,z = z+eta/2*glpdf/lpdf might need go through replace nan process as mentioned in introduction. And the evaluation of convergence is done with the 3rd reward (competitive level). And in addition to the accuracy and evaluation, output the reward lpdf in the last step denote as ld instead of the signal.

B. *Pre-network in Generator*

In the generator block, the aim is to generate adversary signal as the competitive player of the original signal. The baseline is the average of the multi ictal signals. For instance, when generate the adversary player for interictal class(any of 1st ,4th and 5th), we have EEG_G0 = mean(ictal_n_interv([2,3]), 2) while for ictal class(any of 2nd and 3rd), we have EEG_G0 = mean(ictal_n_interv([1,4,5]), 2).

The classification process is the same as in A except for the final reward is denoted as lg and output.

Algorithm II(Block G, continued from block D)
1.get Input (ld, baseline EEG_D0, baseline EEG_G0) from D block
2.generate adversary signal with gamma distribution:
mu = mean(EEG_D0),
temp=gammainc(EEG_G0), order=1
EEG_G= EEG_G0/abs(mu).*gammaincinv(abs(temp)/max(temp));
3.do the same as Block D step 2 according to previous work
4. calculate reward combining D and G block with Hamiltonians only first:
lh = -0.5*mean(ld)-0.5*mean(lg)

C. *Comprehensive Zero-Sum Rewaard Realization with angle, phase, power information.*

Semantics related to game theory uses agent programming language to extend the basic action theory into a domain theory DT = (AT, ST, OT), where ST is the stochastic theory and OT is the optimization theory, defining stochastic actions reward functions and utility functions. Our classifications: ictal and interical are set as the agent o and agent a, which in our example are single agent as utilized in GAN. Although five EEG in all, they are classified individually as independent signals.

1. AT:

In the AT, the primitive actions A and O of agents a and o respectively can be either **single-agent action{φ(a,x1,y1,t1,v1)}over A(resp., {φ(o,x2,y2,t2,v2)}over O)** or **two-agent action{ φ(a,x1,y1,t1,v1), φ(o,x2,y2,t2,v2)}**.

Example(ictal action potential, interictal action potential**):** {ictalFire(o, 1, 20,1,14)}, interictalFire(i, 1, 5,1,10)} stands for the ictal signal fire at time point 1ms, coordinate(1,20) with voltage 14 muV and interictal signal fire at time point 1ms, coordinate(1,5) with voltage 10 muV

2. ST:

In the ST, a set of axioms, stochastic actions are constitute through deterministic actions. We define the deterministic components, n of stochastic action c in situation s and chosen by nature with probability p, denoted as a **predicate stochastic(c,s,n,p).** With the deterministic components n of c in s, we thus have the probability function , denoted as **prob(c, s, n).**

Example(ictal propagate 1 unit along y, interictal propagate 1 unit along x **):** stochastic({ictalFireS(o,1,20,1,14),interictalFireS(a,1,2,1,10)},s,{ictalFireTo(o,1,20,y',t',v1'),interictalFireTo(a,x',2,t',v2')},p) def (x'=2^t'=2,v1'=13^p = 0.5)V(y'=2^t'=2^v2'=9^p=0.5)

Another predicate to check which of the two deterministic components is actually executed is **condStAct**.

Example(Check which of above actions is executed individually for one single player case **):** condStAct({ictalFireS(o,1,20,1,14)},s,{ictalFireTo(o,x',2,t',v2')}), def at(o, x',2,t',v2')^(x'=2^t'=2,v1'=13^p = 0.5), condStAct({interictalFireS(a,1,2,1,10)},s,{interictalFireTo(a,1,20,y',t',v1')}),def at(a,1,20,y',t',v1')^(y'=2^t'=2^v2'=9^p=0.5)

3. OT:

The reward function, utility function and Nash selection functions is the three basic components in OT. A zero sum rewards function associates with competing agents a and o, can be denoted as the reward to agent a: **reward(α, s)** and the reward to agent o: **-reward(α, s)**

Example: reward({ictalFireTo(o,1,20,2,14)}, s)=r, stands for the reward to classify signal as ictal, at coordinate (1,20) and timepoint 2ms with voltage 14 is r;

One important function to evaluate the executability of a program is utility function:**utility (v,pr),** maps a pair of reward and its probability to be executable pr to a real-valued utility**.**

Example: utility(v,pr) = v*pr;

Nash selection functions are the rules which are usually pre-interpreted and to make sure no rewards can be larger than the selected locally at least, they are thus not explicitly axiomatized in domain theory.

In this mix block, the competition process is evaluated with zero-sum rewards of the signals in D and G blocks with respect to their angle, phase and power information.

Feature selection(extended part):

1. (Interictal ^ c1(D Angle range is larger than G Angle range)^ r = -1) V (Ictal ^¬c1(D Angle range is smaller than G Angle range) ^ r = 1)
2. (Interictal ^ c2(D Phase range is larger than G Phase range)^ r = -1) V (Ictal ^¬c2(D Phase range is smaller than G Phase range) ^ r = 1)
3. (Interictal ^ c3(D Phase median is larger than G Phase median)^ r = -1) V (Ictal ^¬c3(D Phase median is smaller than G Phase median) ^ r = 1)
4. (Interictal ^ c4(D Inter ictal power range is larger than G power range)^ r = -1) V (Ictal ^¬c4(D ictal power range is smaller than G power range)^ r = 1)
5. (Interictal ^c5(D Inter ictal power median is smaller than G power median)^ r = -1) V (Ictal ^¬c5(D Inter ictal power median is larger than G power range) ^ r = 1)

Process:

```
Rextd(i)=0,c(i)=ci, k,  i=1,2,3,4,5
  |                          |<-----------k-1-------------| else
  |->Input. lh,  idx in [1, 4, 5]?->c(i)?→Rextd(i)+1→k ==0?
  |                                          |
  |                      lextd(i)=  -log(rexd(i))/sum(log(Rextd(i)))-lh
  |                                                              |
  |-else->--c(i)?--rextd(i)-1→k==0?-->  lextd(i) = lh            |
       |                     | -log(-Rexd(i))/ sum(log(Rextd(i)))  |
       |←-------k-1------else      |                              |
                      mul = mean(lextd)       mul = mean(lextd)
                      c(lextd<mul) = nan      c(lextd>mul) = nan
```

Finally, Select features which are not nan.
(Note that, regression is applied on specific problem and proof that zero sum rewards are all good features see Appendix(Figure 8).[5])

D. LDA based on the distance of the similarity

Suppose we consider the classification with Bayes optimal on the assumption of homoscedasticity, which means the identical value of the covariance of the two classification: sigma_D = sigma_G denoted as Sigma, thus we have the some equal quantities with reagrds with the covariance and cancelled with each other and set the threshold as:

C = w*0.5*(mean(EEG_D) - mean(EEG_G)),

where w =(mean(EEG_D) - mean(EEG_G))/Sigma.

$$\text{EEG\_test} = \begin{cases} \text{ictal}, & \text{if } w * \text{EEG\_test} > c \\ \text{iterictal}, & \text{otherwise} \end{cases}$$

*Alternatively, higher order data can be classified with the QDA kernel, computing the log likelihood ratios:
Criteria = (EEG_test - mean(EEG_G)^T *(EEG_test - mean(EEG_G)/sigma_G + ln|sigma_G| + (EEG_test - mean(EEG_D)^T *(EEG_test - mean(EEG_D)/sigma_D + ln|sigma_D|,
And compare it with the threshold in the kernel.

IV. Experiment and Results analysis

This deep network pre-train the GANs on basis of hamiltonians which iterates 50 epochs each on discriminators and generator(signals see figure 1), which later is regarded as the competing players of the feature matching contrast model[12]. Furthermore, the FOL is applied to generate conditions for feature selection. Due to the limit of the paper length, the reward function, SGD combined regression is not tested with validation before put into LDA model. And the final test on a distance based model is to break the limit on non-expressiveness format. The experiment shows the the computation of 5th epilepsy EEG record which is manually labeled as interictal.

A. Discriminator

After the regular pre-processing(figure 0), the pre-trained network starts training with hamiltonians evaluating the similarity of the white noise mixed original EEG_D. The descriptive statistics of the IMFs after EMD(table 3) shows mainly the intensity of the interictal and ictals. As we can also see (from the figure 2), the 5th IMF gives most instantaneous information of the signal and there is the most intensive spike series existing from 10-15 of this signal. And the instantaneous energy has largely loss stocked on any mono IMF. And thus, instead of the mean, we store the median value of each quantities avoiding some bias introduced by the pre-computation(and in accordance with the median values all around 0 with regards to voltage and the minimum is although negative but is close to zero as well). And although the Hamiltonians give the measure of the signal dynamically with the position and velocity related quantities, it is not achieved explicitly which is based on computation of potential in our paper. And it is shown to be all positive for both the two Hamiltons. Note that to compare the instant measurement on EEG with some averaged reference, the fft transformed value is also stored. Actually, the difference of the energy is although not small but not on out of tolerance.(two-sided squared power max : 19301239.7001 while maximum instantaneous energy is: 739.041156, 35 fold).

The final prediction accuracy(figure 3) for the whole block is quite high as 1. And dig into some details, the likely hood is not highest with batchsize 5 while lowest while batchsize 1, in consistent with the log ratio of p,q, giving Ad the specificity is generally low but quite high on the 4th and similarly for miss rate which is good as 0 at first three but goes to 0.5 for the last two, showing certain robustness. And the accuracy for each batch is also rising from around 0.4 to 1 with std also getting smaller giving good performance. Finally, with regard to the reward evaluation, five value although does not show convergence, but are all close to zero which is also accepted as giving non-divergent behave but with some uncertainty still.

B. Generator

Since the aim of the generator is to genrate significantly different signal competing with original signal, the properties is indeed significantly different for both dynamics and average quantities according to two-sampled ttest (p = 1.8455e-6 with h=1). As the generator gives signal with some randomization as the distribution is gamma distribution with magnitude and baseline(mean of the same multi channel of epilepsies) value are certain. The experimented one has the most information in the first IMF.(See figure 4), and in the first 0-5 time iterval. And the descriptive statistics of the voltage only is quite similar to the original signal if only consider median(which cann also be seen from the residue in figure 1). However, the angle, phase properties are totally different from the original signal. For instance, the instantaneous frequency median is from 0.999927 changed to 2.359427 and phase median from 2.86 to 3.25. which also leads to the significance of the different hamiltonians.

Similar to the D block, we also do the accuracy computation on 8 different quantities and one reward evaluation on the 5 batch sized experiment here. Not surprisingly, as the randomized information is quite marge, the final total prediction(figure 5) is slightly lower(0.8) with mistakenly prediction on the last signal. To further evaluate, the likelihood is of quite similar trend although being negative as they are adversary signals. And the log likelihood ratio are higher for the later 4 signals than the first one. For those robustness related quantities, it has better sensitivity(all zero but middle one being 0.5) than the decoder both but poorer specificity on first two generated eplipsy.As for the accuracy, it is all but the third not being 1.

C. LDA after optimized features

As introduced detailedly in the chapter 2, the feature selection is done with linear regression optimizing the zero sum reward function based on both reward from the pre-training network consist both D-block and G_block along with the LDA model. As the LDA is a distance-based on kernel with regards to covariance and the center of the two classifications and the computation of the theshold and criteria is not hard to understand, it is commonly used on machine learning and data mining, especially fast with data has feature selected. Here, we use the network with data filtered (24 features left: 'IMFDMedInI', 'IMFDMINInI','IMFDMAXInI', 'IMFEDMedInI,'IMFEDMINInI', 'IMFEDMAXInI', 'IMFFDMedInI', 'IMFFDMINInI', 'IMFFDMAXInI','P_dMed', 'P_dMax', 'Q_dMed', 'ANGLEDMIN', 'ANGLEDMAX', 'PHASEDMed', 'PHASEDMAX', 'POWERDMed', 'POWERDMIN', 'POWERDMAX').

The final LDA (FIGURE 6). is ietrated on 30 epochs, optimizing the the objective function costs 0.71206 in the 26.599 elapsed time, during the 30 epochs, the observed objective function value is 0.85 with best oberved feasible point(delta = 0.086169, Gamma = 0.84728) while in the last epoch, we reach the estimated objective function (FIGURE 7). at 0.84903 using 0.022317 with best observed feasible point(alpha =1.3288e-05, gamma = 0.0722317).

V. DISCUSSION

As the technology is accelerating the growth of the information, machine learning and data mining out-stands some traditional models advances in autonomously compute and solve some systems without knowing the exact mechanism of it as a black box. And to cope with the exponentially increasing data and network, the logic is founded on a stochastic notion of different problems, including reasoning about goals and intentions. Related application build models programming the agents combing game theory and domain theory. Like how the knowledge base is well applied on the semantics web, the game theory also is applied commonly on decision-making and optimizing problems. As the traditional semantic models origin from FOL are usually limited to non-expressive knowledge representation, the transition of the database to knowledge bases and more specifically, the information content based contrast model works well on the common infrastructure as an inductive learning method. The autonomous agents not only behaves as as independent decision-makers, the relationships between them and recent progress towards specific formalism for expressing social concepts using the domain theory consists of the basic action theory, stochastic theory and optimization theory.

This paper utilize the GANs, a commonly used image process model on the EEG data to analogously solve a binary classification problem through semi-supervision. In the supervised part, i.e. the discriminator, the network is pre-trained with comparing the features, mainly with energy-based rewards, Hamiltonians. Similarly, after generating adversary signal with baseline averaged by origin data, using gammainc distribution convolution, considering both the spatial and temporal information, it is agian trained with Hamiltonians and further with more features linear regressionly selected combining stochastic degradation(SGD) search. Note that such a computation requires large cost on physical memory and time, thus we simplize it with variance inference on the data generation part, avoid generate it in every epoch inferred by previous while only train it dynamically heuristically, which is also inspired by the origin ontology idea of semantics web multi-recourses open world problem with the stochastic model composed with some deterministic parts.

Finally, after the feature selection based on zero-sum reward, combining description logics(DL), the LDA kernel is applied to test the pre-trained network on a higher level, computing similarity with the co-variance, classifying the EEG into ictal and inter-ictal with the derived criteria from QDA aiming at maximizing inter-variance along with minimizing intra-variance. And the remark some partial observability and strategic reasoning ability[14], this assembling model posses some advances of the stochastic multi-agents system as well as put forward the GAN model on genral data and assembeled with the LDA model.

The future interest and direction will be combine energy -based rewards with some explicit expressive model, find some analytical solution instead of approximation only and to again develop the model into a complete stochastic game is also of my interest.

ACKNOWLEDGMENT

Gratitude to my bachelor thesis and master thesis supervisors as well as the current lab supervisor as the work and study in the labs help me a lot on the article.

REFERENCES

[1] C. Shannon A Mathematical Theory of Communication, Bell System Technical Journal, 1948

[2] Yuhang Song, Andrzej Wojcicki, Thomas Lukasiewicz, Jianyi Wang, Abi Aryan, Zhenghua Xu, Mai Xu, Zihan Ding, Lianlong Wu, Arena: A General Evaluation Platform and Building Toolkit for Multi-Agent Intelligence

[3] Mayur Saxena, Aidan Gomez, Yarin Gal, 'Location Conditional Image Generation using Generative Adversarial Networks', neurips2019 conference

[4] Vito Walter anelli, Renato De Leone, Tommaso Di Noia, Thomas Lukasiewicz, Combining RDF and SPARQL with CP-theories to reason about preferences in a Linked Sata setting

[5] F. Baader, D。 Calvanese, D. McGuinness, D. Nardi, and P.Patel-Schindler, Editors. The Description Logic Handbook, Cambridge University Press, Cambridge, 2003.

[6] Lambert, Max; Engroff, Andrew; Dyer, Matt; Byer, Ben, 'Empirical Model Decomposition.'

[7] Chen, Yangkang; Ma, Jitao (May-June 2014). 'Random noise attenuation by f-x empirical-mode decomposition presictive filtering.' Geophysics.79(3):V81-V91. Doi: 10.1190/GEO2013-0080.1

[8] Blenca Rodriguez, Adam McCarthy, Ana Mincholé, University of Oxford, Variational Inferenceover Non-differentiable cardiac Simulations using Bayesian Optimization Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[9] Oana-Maria Camburu, Brendan Shillingford, Pasquale Minervini, Thomas Lukasiewicz, Phil Blunsom, Make Up Your Mind! Adversarial Generation of Inconsistent Natural Language Explanations Electronic Publication: Digital Object Identifiers .

[10]] Anthony L. Caterini, Arnaud Doucet, Dino Sejdinovic , Hamiltonian Variational Auto-Encoder European Space

[11] A. Kito, Y. Mizumachi, K. Sato, Y. Matsuoka, "Emergent Design System Using Computer-Human Interactions and Serendipity" The Sixth International Conference on Advances in Computer-Human Interactions (ACHI 2013) IARIA, Feb. 2013, pp. 7-12, ISSN: 2308-4138, ISBN: 978-1-61208-250-9

[12] S. Brandt, R. Kusters and A.-Y. Turhan, Approximation and difference in description logitics, in Processings of the 8th International Conference on Principles of Knowledge Representation and Reasoning, KR02, D. Fensel, F. Giunchiglia, D. McGuinness, and M.-A. Williams (Eds.), Morgan Kaufmann, San Francisco, 2002, pp.203-214.

[13] B. Scholkopf, and A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[14] M.Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In G. GGopalakrishnan and S. Qadeer, editors, Proc. 23rd International Conference on Computure Aided Verification (CAV'11), volume 6806 of LNCS, pages 585-591. Springer,2011

[15] Nicola Fanizzi, Claudia d'Amato, Floriana Esposito, Induction of Optimal Semi-distances for Individuals based on Feature Sets

[16] Qin He ,Jukka Peltola, Susanna Narkilahi, Jari Hyttinen, Sampsa Pursiainen, XiaochuanPan, Rubin Wang

Ictal classification with Hamilton Monte Carlo Markov Chain Based on intrinsic mode function and empirical mode decomposition.
https://drive.google.com/file/d/1zfRj3j8IeLhCRKAR0k_2JchbTB9zInYs/view?usp=sharing

Appendix

1)Ictals annotated manually by physician:

ictals	duration
ictal1	55s-80s
ictal2	167s-183s
ictal3	216s-343s
ictal4	383s-410s
ictal5	547s-573s

TABLE I *Ictals annotation*

2) Time-frequency coupling:

Specific bands Hz(coupled windows):	time bin
delta = [0.5, 3]	2, 3
theta = [4, 7]	5,6,7
alpha = [8, 12]	9,10,11,12,13
mu = [7.5, 12.5]	14,15,16
SMR = [12.5, 15.5]	17-31
beta = [16, 31]	33-100
gamma = [32, 100]	71-2000
HF = 70 ripple = [80, 250]	81-250
fastripple= 251	252-2000

TABLE II COUPLING BANDS AND TIME BIN

Figure 8. Proof

3) Crit: 0.1852, max dev:3.8415
BetweenSigma = [-1.8143, -2.1547; -2.1547; 1.8143];
4) After LDA analysis:
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 26.599 seconds.
Total objective function evaluation time: 0.71206
Best observed feasible point:

Delta	Gamma	Delta	Gamma
0.086169	0.84728	1.3288e-05	0.31336

Observed objective function value = 0.5
Estimated objective function value = 0.49035
Function evaluation time = 0.032425
Best estimated feasible point = 0.849035
Estimated function evaluation time = 0.022317

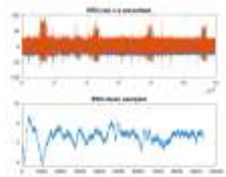

Figure 0, Preprocess

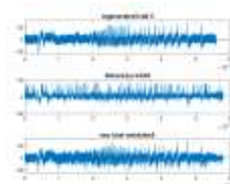

Figure 1. EEG D. EEG G.

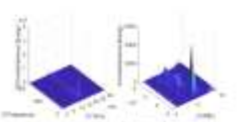

Figure 2. IMF, instantaneous quantities(D block)

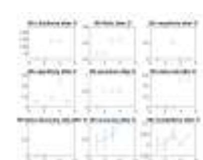

Figure 3. ACCURACY_D

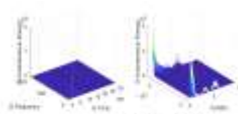

Figure 4. IMF, instantaneous quantities(G block)

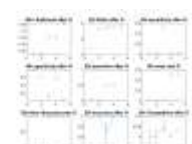

Figure 5, ACCURACY_G

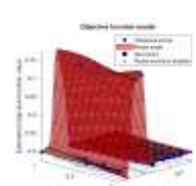

Figure 6. LDA optimize

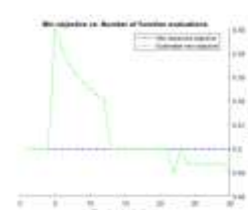

Figure 7. LDA minimize