35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Automatic Optimization of Parameters for Seizure Detection Systems*

P. Dollfuß, M.M. Hartmann, A. Skupch, F. Fürbaß and T. Kluge

Abstract—**A parameter optimization method for an automatic seizure detection algorithm using the Nelder Mead algorithm is presented. A suitable cost function for joint optimization of sensitivity and false alarm rate is proposed. The optimization is done using EEG datasets from 23 patients and validated on datasets from another 23 patients. The resulting sensitivity was 82.3% with a false alarm rate of 0.24 FA/h. This is a reduction of the false alarm rate by 1.58 FA/h with an acceptable loss of sensitivity of 4.3%.**

I. Background and Introduction

A. *Epilepsy & Automatic Seizure Detection*

Approximately one percent of the world's population suffers from epilepsy, a chronic dysfunction of the brain that is characterized by recurrent unprovoked and unpredictable seizures caused by an excessive discharge of groups of neurons. Long-term electroencephalogram (EEG) recordings over several days are the corner stone for the presurgical workup for these patients. These recordings and their analysis are extremely time consuming and require medical experts. Online seizure detection systems therefore are a great benefit leading to improved safety for the patients and reduced costs.

AIT has developed a high performance epilepsy seizure detection (ESD) system for long-term EEG monitoring [1]. The algorithm is based on a frequency domain method called Periodic Waveform Analysis (PWA) and the time domain analysis of epileptiform sequences (EWS) [2]. The system convinces with a very good detection performance and no need of patient dependent parameter adjustment. Nevertheless parameters are used in the decision making process as thresholds limits. Since the parameters have an impact on the performance of the system, adapting these parameters is quite critical and essential. Even if some prior knowledge on the thresholds (scatter plots etc.) is available, setting the system parameters manually is hardly possible. Therefore an automatic parameter optimization is used, presented in this paper.

B. *Optimization Algorithms*

An optimization algorithm aimed at finding the optimal set of parameters leading to an optimum performance of a system. The meaning of optimality has to be defined by an appropriate cost function. There are a large number of different optimization techniques. To understand the choice of the proper optimization algorithm for a seizure detection system, a small overview on the most important methods and the requirements is given.

Requirements: The goal of an optimization algorithm is to find a global optimum set of parameters within finite time. In fact most of the optimization algorithms can just find estimates of local optima. The more detailed requirements based on system or parameter restrictions are even more essential. Some examples are listed below:

- the system is linear/nonlinear
- discrete parameters are allowed only
- no gradient is available
- no model is available

In the following part the main optimization methods are described:

Analytic methods try to solve the optimization problem by finding a direct analytical solution based on the system equations.

Iterative methods try to estimate better parameters on the base of previous results and the associated parameters. Thus the methods converge iteratively into an optimum. Representative methods are the pattern search methods or Newton's method [3].

Heuristic methods use trials or heuristic changes in the parameters set to reach an optimum result. Most important heuristic methods are the nature-inspired algorithms like the genetic algorithms [4] or the bee algorithm [5].

In the next section, the optimization method used for a seizure detection system with its special needs in the optimization process is discussed.

II. Methods and Data

A. *Cost function*

The performance of an ESD is typically defined by two statistical measures, the sensitivity (also known as hit rate) and the false alarm rate [6]. We defined the sensitivity and the false alarm rate as follows:

Sensitivity and False alarm rate calculation: the sensitivity is calculated as the ratio of true positives and the total number of recorded seizures. The derivation of the false alarm rate is more complex because seizure alerts are clustered to sub-alerts with duration of 30 seconds. Each sub-alert that does not intersect with a true seizure marker (basic

*This work was supported in part by HIGH PROFILE. The HIGH PROFILE project has received funding from the ARTEMIS Joint Undertaking under grant agreement n° 269356.

Peter Dollfuß, Manfred M. Hartmann, Ana Skupch, Franz Fürbaß and Tilmann Kluge are with the Austrian Institute of Technology, Donau-City-Strasse 1, 1220 Vienna, Austria (phone: +43 50550 4227; e-mail: Peter.Dollfuss@ait.ac.at).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

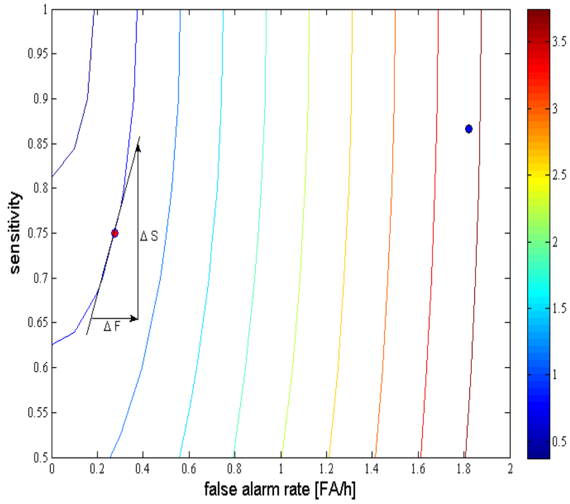

Figure 1: Plot of the cost function used with this optimization procedure. The left dot shows the operating point, used to design the cost function (see II.A). On the right hand the real starting point of the process can be seen.

truth) is regarded as a false alarm. The calculations of the performance measures are done patient-wise and can be found in more detail in [1]. The overall performance across all patients is done by averaging the patient-wise sensitivities and false alarm rates.

Cost function design: By dealing with two measures at the same time the optimization problem becomes a multi-criteria problem. Instead of using Pareto-optimization approaches [7] (also known as multi objective optimization) the partial-cost judgment is done within a cost function which depends on both, sensitivity and false alarm rate.

Generally the main objective of the optimization, especially by using the Nelder Mead algorithm, is to minimize a cost function by adjusting the parameters. Hence the optimization strategy or the desirable performance is defined by the cost function. In our case the desired performance values ideally are 100% sensitivity with no false alarms.

However the choice of a suitable cost function is not obvious. Here we choose a function depending on the false alarm rate F and the sensitivity S, which yields equal cost values on elliptic contour planes centred around 100% sensitivity and 0 false alarms per hour:

$$f(F,S) = \sqrt{\alpha^2 F^2 + (1-S)^2}. \tag{1}$$

To design the shaping parameter α, first the partial derivatives with respect to the false alarm rate $\frac{\partial f(F,S)}{\partial F}$ and the sensitivity $\frac{\partial f(F,S)}{\partial S}$ are determined.

With these derivatives, the slope g of the contour plane in a chosen operating point (F_{op}, S_{op}) can be calculated as follows:

$$g(F_{op}, S_{op}) = \frac{\frac{\partial f(F_{op}, S_{op})}{\partial S}}{\frac{\partial f(F_{op}, S_{op})}{\partial F}} = \frac{-1 + S_{op}}{F_{op}\alpha^2}. \tag{2}$$

Now, α will be calculated such that $g(F_{op}, S_{op})$ in the chosen point of operation is equal to a slope $\frac{\Delta S}{\Delta F}$, where ΔS is the gain in sensitivity while accepting an increase of the false alarm rate by ΔF:

$$g(F_{op}, S_{op}) = \frac{-1 + S_{op}}{F_{op}\alpha^2} = \frac{\Delta S}{\Delta F}. \tag{3}$$

Equation (3) can be rewritten as:

$$\alpha = \sqrt{\frac{\Delta F}{\Delta S}} \sqrt{\frac{S_{op} - 1}{F_{op}}}. \tag{4}$$

For our optimization the operating point (F_{op}, S_{op}) was chosen to be at 75% sensitivity and 0.3 FA/h. The gain in sensitivity ΔS and the increase of false alarm rate ΔF were set to 2% and 0.2 FA/h.

The resulting cost function, with the operating point and the starting point of the process is shown in Figure 1. By considering one particular contour of the cost function it becomes clear that performance values lying on the contour lead to similar cost values.

B. Optimization algorithm

As described in subsection I.B, there are different approaches for optimization algorithms. The choice of the proper algorithm is essential and strongly depends on the system and parameter properties. To explain the final choice of the algorithm the most important properties are listed below:

- the parameter space is multidimensional
- the result space is discrete, non-smooth and not differentiable

Choice of the algorithm: The analytic methods rely mainly on linear problems and on system models. Due to this fact the methods are out of the question. The heuristic methods constitute a possible approach, but cannot guarantee a solution. Concerning the listed properties the iterative methods have been selected. The methods convince with their simple principles and the possibility to adapt on special constraints. Among this class of algorithms the downhill simplex method or also known as Nelder Mead algorithm was chosen [8]. The major criterion for the choice of this algorithm has been the fact that the algorithm doesn't need derivative values. Additionally the algorithm is easy to implement and the most common technique used in practice.

C. Data and initial parameter values

To avoid overfitting on one particular dataset a strategy from the field of machine learning is used: Two independent datasets are used to prove the relevance of the optimization results. The first data set is used as a training set thus is directly involved in the optimization process. The second set, the validation set, is not involved into the optimization

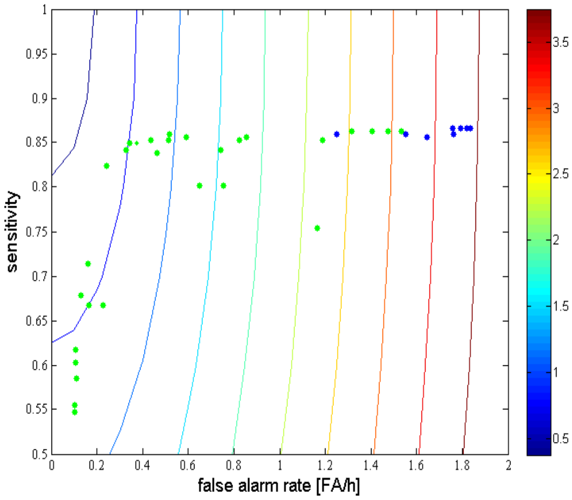

Figure 2: Plot of the particular results during the optimization process. To distinguish the starting trials from the real optimization results, the starting trials are colored in blue. The color bar shows the actual cost values thus the costs of the particular results. Solid lines represent sensitivity/false alarm rate pairs of equal cost function.

process but also uses the parameters resulting from that process. To prove "non-overfitting" and to show the effect on the parameter changes both datasets have been fed by the starting parameter set (the set before the optimization) and the optimized parameters set.

Both datasets consist of 23 EEG recordings from 23 different patients with on average four epileptic seizures in each dataset. The recordings of the two dataset groups have been done in different Epilepsy Monitoring Unit.

Basic truth: In order to analyze the performance of a seizure detection system, annotations of seizures that are visible in EEG are required. With our system the annotations are done offline by EEG technicians [1].

Initial algorithm parameter set: The parameters to be optimized are 12 threshold values used in the final decision making process (alert or no alert) of the algorithm. The number of 12 parameters results from two different feature (EWS [2] and PWA [1], mentioned in I.A), split into six frequency bands, involved in the decision making process. For that reason the parameter space results in 12 dimensions.

Initial values have been determined manually during the development of the feature extraction algorithm: For a small set of data including a number of different seizures, parameters where chosen such that a good discrimination of seizure- and non-seizure activity was achieved. Note that this is only feasible for relatively small data sets. The automatic optimization was executed on a substantially larger amount of data.

Beside the starting values the values of the first parameter trials (start trials) during the starting procedure of the Nelder Mead algorithm have to be chosen [8]. Those values are essential for the convergence and especially for the convergence speed. During the development knowledge about the particular influence of the threshold parameter changes on the system performance has been gained. Thus with this prior knowledge the actual values for the trials are set on the way to deliver significant cost values changes within the starting procedure.

Optimization algorithm parameters: The parameters of the optimization algorithm are set to recommend values given in [8]. If the optimization is driven automatic the breaking condition or the maximum number of Iterations has to be set individually.

III. RESULTS

The optimization procedure has been stopped manually after 42 iterations due to the fact that the algorithm had already converged to a satisfactory solution. The effort of the optimization through these steps has been roughly two weeks, dependent on the loads of the computers (six "state of the art computers" with four cores).

Figure 2 shows the results of the optimization procedure. Each green dot represents the results of one iteration step, starting values are represented by blue dots. The best performance result related to the lowest cost value results in 82.3% sensitivity and a false alarm rate of 0.24 FA/h. Compared to the starting performance values (cf. row 1 in Table 1) a strong reduction in false alarm rate of 1.58 FA/h is achieved. Given the huge reduction in false alarm rate, the loss of sensitivity of 4.3% sensitivity is acceptable.

It can be seen in Figure 2, that false alarm rates below 0.1 FA/h can hardly be achieved, and that sensitivity values above 0.86% are also hardly achievable. Most of the parameter sets tested by the optimization algorithm lie within a narrow band that seems to describe a receiver operating curve (ROC [6]) for this seizure detection algorithm. The choice of the given cost function (II.A Equation (3)) effectively favours operating points with high sensitivity and low false alarm rate (left upper corner in Figure 2). It is easy to conceive, how another choice of the shaping parameter α would have resulted in a different location of the resulting "optimum" operation point: a smaller value for α results in an optimum point "on the left" on the ROC, i.e., with lower sensitivity and false alarm rate, a larger value for α results in an optimum point "on the right" on the ROC, i.e., with higher sensitivity and false alarm rate.

To prove the statistical significance (and thus the absence of over-fitting) we used separate validation- and training data sets (see II.C). The results for these separate data sets are shown in Table 1: For both, training and validation data sets, performance values were initially calculated for the starting parameter values. Similar results were achieved for both data sets. Optimization was accomplished using the training data set and then the performance was evaluated for both data sets using the optimized parameters.

It can be seen in the last rows of Table 1, that the sensitivity was 82.3% for the training set and 80.8% for the validation data set, and the false alarm rate was 0.24 FA/h for the training set and 0.26 FA/h for the validation data set. One can see that performance values for the validation data set are very similar to those for the training data set. This shows that the optimization effect carries over also to independent data and is not an effect of data over-fitting.

TABLE 1

	Datasets	FAR	SENS
Starting parameter set	Training set	1,82	86,6 %
	Validation set	1,96	88,4 %
Opimized parameter set	Training set	0.24	82,3 %
	Validation set	0.26	80,8 %

IV. SUMMARY

A parameter optimization method for an automatic seizure detection algorithm has been presented. Considering the system properties, i.e., a non-smooth result space and a multidimensional parameters space, the well-known method of Nelder Mead was chosen. This method requires the definition of a suitable cost function. Therefore, the performance measures characterizing a seizure detection system had to be combined to a single cost function value.

A method for choosing an appropriate shaping parameter has been presented. For the optimization we used a set EEG datasets from 23 patients, each containing four seizures on average. To avoid over-fitting a validation dataset from another 23 patients was chosen.

The optimization process was stopped after 42 iterations due to an already satisfactory performance. The resulting sensitivity was 82.3% with a false alarm rate of 0.24 FA/h. Compared to the initial values the false alarm rate could be reduced by 1.58 FA/h with an acceptable loss of sensitivity of 4.3%. These performance values could also be reproduced with the validation data set: here the sensitivity with optimized parameters was 80.8% at a false alarm rate of 0.26 FA/h. This shows that the gain was not achieved due to over-fitting, but could also be achieved with the validation set.

REFERENCES

[1] M. M. Hartmann, F. Fürbass, H. Perko, A. Skupch, K. Lackmayer, C. Baumgartner, und T. Kluge, „EpiScan: Online seizure detection for epilepsy monitoring units“, in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, S. 6096 –6099.

[2] F. Fürbass, M. Hartmann, H. Perko, A. Skupch, P. Dollfuss, G. Gritsch, C. Baumgartner, und T. Kluge, „Combining time series and frequency domain analysis for a automatic seizure detection“, in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, S. 1020 –1023.

[3] R. Hooke und T. A. Jeeves, „“ Direct Search” Solution of Numerical and Statistical Problems“, *J. ACM*, Bd. 8, Nr. 2, S. 212–229, Apr. 1961.

[4] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.

[5] D. Karaboga und B. Basturk, „A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm“, *J Glob Optim*, Bd. 39, Nr. 3, S. 459–471, Nov. 2007.

[6] T. Fawcett, „An introduction to ROC analysis“, *Pattern Recognition Letters*, Bd. 27, Nr. 8, S. 861–874, Juni 2006.

[7] M. Ehrgott, *Multicriteria Optimization*. Springer, 2005.

[8] J. A. Nelder und R. Mead, „A Simplex Method for Function Minimization“, *The Computer Journal*, Bd. 7, Nr. 4, S. 308 –313, Jan. 1965.