RELIABILITY IN PREPROCESSING - BAYES RULES SIESTA

Sykacek P.[1], Roberts S. J.[2], Rezek I.[2], Flexer A.[1], Dorffner G.[1]
[1] Austrian Research Institute for Artificial Intelligence,
Schottengasse 3, A-1010 Vienna, Austria
[2] Robotics Research Group, Department of Engineering Science,
University of Oxford, Oxford OX1 6PJ, UK

peter@ai.univie.ac.at

Abstract: The SIESTA project aims at defining a new description of human sleep. As such, the SIESTA sleep analyzer is a diagnostic tool applied to biosignals of humans. Although the risk that wrong decisions harm people is low, it is still there. Hence introducing a reliability measure that flags decisions that are probably wrong is a major aim of the project. This paper introduces a method for preprocessing within the Bayesian framework. We show that Bayesian beliefs can be used to flag segments where reliable decisions about the sleeping subject are not possible.

Introduction

We decided to embed preprocessing in a Bayesian framework because it provides means to solve all problems that emerge during estimation of models from data [2]:

- Inference of model coefficients.
- Treatment of nuisance parameters like noise levels.
- Reporting beliefs as necessary for model order estimation and sensor fusion of adjacent segments.

Preprocessing is done by auto-regressive (AR)-models. We use a lattice filter representation of an AR-process and infer the a-posteriori distribution over model coefficients - the so called reflection coefficients. Reporting our beliefs in a model is used for model order estimation and as reliability measure.

Methods

In [1] the lattice filter representation is related to AR-coefficients by the so called Levinson algorithm:

$$\begin{aligned} a_k^{m+1} &= a_k^m + \rho_m a_{m-k+1}^m \\ a_{m+1}^{m+1} &= \rho_m \end{aligned} \tag{1}$$

The superscripts used in (1) denote the model order, the subscripts are the indices of the AR-coefficients and ρ_m is the m-th order reflection coefficient. We assume a uniform "stability" prior over ρ_m in the interval $[-1, 1]$ and a uninformative Jeffrey's prior over the noise level. Some calculations lead to:

$$\begin{aligned} p(\rho_m|\mathcal{D}) \propto\ & \sqrt{2\pi}s 0.5\pi^{-\frac{N}{2}}\Gamma(\frac{N}{2}) \\ & \left((\underline{\epsilon}^m + \hat{\rho}_m \underline{r}^m)^T (\underline{\epsilon}^m + \hat{\rho}_m \underline{r}^m)\right)^{-\frac{N}{2}} \\ & \underbrace{\frac{1}{\sqrt{2\pi}s}\exp\left(-\frac{1}{2s^2}(\rho_m - \hat{\rho}_m)^2\right)}_{\text{normalized Gaussian}}, \end{aligned} \tag{2}$$

as an expression of the a-posteriori distribution of the m-th order reflection coefficient with:

$$s^2 = \frac{1 - (\hat{\rho}^m)^2}{(N-1)},$$

as variance. In (2) $\underline{\epsilon}^m$ and $\underline{r}^m$ are the forward prediction error and the backward prediction error at lag -1 respectively. The most probable reflection coefficient, $\hat{\rho}_m$, is given as a least squares solution.

The evidence factor, $p(\mathcal{D}|I_m)$, in (2) displayed in front of the Gaussian, can be used for reporting our belief in the model. However, we can only compare against another model. In our case an appropriate choice to compare against is a model that does not use the coefficient ρ_m and declares the signal as noise only. The evidence of this second explanation of our data is:

$$p(\mathcal{D}|\bar{I_m}) = \pi^{-\frac{N}{2}}\Gamma(\frac{N}{2})\underline{\epsilon}^{mT}\underline{\epsilon}^m. \tag{3}$$

If we use a uniform prior over models, we get

$$P(I_m|\mathcal{D}) = \frac{p(\mathcal{D}|I_m)}{p(\mathcal{D}|I_m) + p(\mathcal{D}|\bar{I_m})}. \tag{4}$$

as a-posteriori probability of that model. Contrary to the common use of a-posteriori probabilities of models, where the m probabilities sum up to 1, we have

a slightly different situation here: Each of the probabilities $P(I_m|\mathcal{D})$ is between 0 and 1 and the optimal model order is the largest index m, where $P(I_m|\mathcal{D})$ is above 0.5.

Assuming that a reflection coefficient is not needed to build a model of the data is equivalent to assuming that the corresponding order m does not contain any information. Hence comparing $P(I_m|\mathcal{D})$ with $P(\hat{I_m}|\mathcal{D})$ means to ask how much belief we have in the hypothesis that a particular data segment contains any information. We therefore argue that $P(I_m|\mathcal{D})$ is also the appropriate belief of the reflection coefficient of the corresponding data segment.

Experiments

The experiments reported here aim to show that the posterior probability $P(I_m|\mathcal{D})$ is indeed a measure that captures the reliability of coefficients extracted from some data. We will try to flag segments marked as artifactual by experts. However, we do not argue that the proposed measure mimics the expert opinion about artifacts. It only works if the data contaminated by artifacts contain almost no information. Some artifacts, e.g. ECG, and small contaminations in general will not lead to lower reliability. Hence recognizing such artifacts is not possible.

The experiments are based on data extracted from different all-night recordings from the SIESTA database. The data were scored by human experts on a one second basis as either contaminated by some artifact, where different artifacts have been considered separately, or as being clean. All together we have 213664 clean segments and 115736 segments marked as artifactual. As our recordings are sampled at different rates, we had to resample[1] to a common frequency of 100 Hz. After resampling, we used two seconds windows and an offset of one second to estimate three reflection coefficients and the corresponding model probabilities.

In order to assess the correlation of our reliability measure with the experts opinion on artifacts, we perform two tests. In a first experiment we try to flag so called movement artifacts, where we have to estimate a rejection threshold. In general we want to have only few false positives, therefore we aim at a specifity of 0.99. The data used to set the threshold were excluded from any further study. Accumulating results from 7 different subjects, we get a sensitivity of 0.298 and a specifity of 0.974. In table 1 we see that there is some inter-subject variation.

A problem of the first experiment is that the reliability measure indicates some problem with a particular data segment. However, it is not designed to separate different artifacts. Hence in a second experiment we looked whether we can find a correlation between low reliability and segments contaminated with *some* artifact. The critical point, where we would better suggest *not* to use a particular model, is at probability 0.5. Using this threshold, we find 2321 segments marked as contaminated and only 504 clean segments. Hence both experiments suggest that there is a correlation between artifacts and our reliability measure. However there are some differences between expert opinions and our reliability measure: Even if we are restrictive and request that *all three* lattice stages have to be extremely implausible (we request that $\forall m$ of a segment $P(I_m|\mathcal{D}) < 0.07$)[2], we still get 3 samples marked as "clean" by experts.

Table 1: Sensitivities and specifities

sens.	0.26	0.13	0.37	0.59	0.21	0.27	0.27
spec.	0.95	0.99	0.99	0.99	0.90	0.99	0.99

Conclusion

We have shown in this paper that preprocessing within a Bayesian framework has two major advantages: it allows to perform model selection and it reports beliefs that can be used for sensor fusion in later stages. To our knowledge, a lattice filter has not been treated within a Bayesian framework so far. The benefits of doing so have been demonstrated by two experiments with a large database of expert labeled data.

Acknowledgements

This work has been done within the project SIESTA, funded by the EC Dg. XII grant BMH4-CT97-2040. The authors want to express gratitude to the clinical partners involved in artifact processing for providing the labels. We further acknowledge that we used code provided by A. Schlögl to load data into MatLab.

References

[1] L. Ljung. *System Identification, Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey, 1999.

[2] J. J. K. Ó Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, New York, 1995.

[1] In these experiments we used the MATLAB signal processing function "resample".

[2] Note that this is equivalent to suggesting that the data are pure noise.