Expert Systems with Applications 39 (2012) 9340–9347

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Classification of newborn EEG maturity with Bayesian averaging over decision trees

V. Schetinin *, L. Jakaite

Department of Computer Science and Technology, University of Bedfordshire, Luton LU1 3JU, UK

ARTICLE INFO

Keywords:
EEG
Brain maturity
Bayesian model averaging
Markov Chain Monte Carlo

ABSTRACT

EEG experts can assess a newborn's brain maturity by visual analysis of age-related patterns in sleep EEG. It is highly desirable to make the results of assessment most accurate and reliable. However, the expert analysis is limited in capability to provide the estimate of uncertainty in assessments. Bayesian inference has been shown providing the most accurate estimates of uncertainty by using Markov Chain Monte Carlo (MCMC) integration over the posterior distribution. The use of MCMC enables to approximate the desired distribution by sampling the areas of interests in which the density of distribution is high. In practice, the posterior distribution can be multimodal, and so that the existing MCMC techniques cannot provide the proportional sampling from the areas of interest. The lack of prior information makes MCMC integration more difficult when a model parameter space is large and cannot be explored in detail within a reasonable time. In particular, the lack of information about EEG feature importance can affect the results of Bayesian assessment of EEG maturity. In this paper we explore how the posterior information about EEG feature importance can be used to reduce a negative influence of disproportional sampling on the results of Bayesian assessment. We found that the MCMC integration tends to oversample the areas in which a model parameter space includes one or more features, the importance of which counted in terms of their posterior use is low. Using this finding, we proposed to cure the results of MCMC integration and then described the results of testing the proposed method on a set of sleep EEG recordings.

1. Introduction

Early diagnosis of abnormal newborn brain development is a challenging problem in the developmental neurology and clinical neonatology. Experts attempt to assess brain maturity by visual analysis of age-related patterns in electroencephalograms (EEG) recorded from sleeping newborns (Holthausen, Breidbach, Scheidt, & Frenzel, 2000; Tharp, 1990). The analysis can take hours of expert work to confidently interpret sleep EEG, as the age-related patterns widely vary during sleep hours as well as between patients, and there are no regular rules for interpretation of these patterns (Cooper, Binnie, & Schaw, 2003). There are neurological evidences that the post-conceptional ages (PCA) of healthy newborns normally match their EEG-estimated ages. In cases when the mismatch is observed during two and more weeks, the newborn's brain development is most likely abnormal (Scher, 1997). Thus, the mismatch between PCA and EEG-estimated ages can alert about abnormal brain development.

In the first publications on EEG assessment of newborn brain development (Parmelee et al., 1968), the experts have visually analysed 47 EEG recordings made in 11 PCA groups between 39 and 43 weeks. The experts have found 10 maturity-related EEG patterns. Then the experts have estimated the PCA of each EEG recording by counting the distribution of the maturity-related patterns. The expert estimates have exactly matched the stated PCA in 27.6% of cases. In 59.5% the matches were within ±1 week, and 77.5% of cases were found matching within ±2 weeks.

In later publications, it has been attempted to learn brain development models from sleep EEG data recorded from newborns whose maturation was preliminary estimated by experts. In Scher, Steppe, and Banks (1996), the regression models have been applied to mapping the brain maturity into EEG index. In Crowell, Kapuniai, and Jones (1978) and Schetinin and Schult (2005), the classification models have been used for distinguishing the maturity levels, at least, for one normal and one abnormal levels of brain development.

The above attempts were aimed at learning a single model providing the maximum likelihood on given EEG data. However such models cannot ensure the maximum accuracy when the likelihood distribution is affected by noise and its shape is multimodal. Besides, the model selection methodology cannot provide estimates of a full posterior distribution which is required for accurate assessment of the uncertainty in model outcomes.

In contrast, Bayesian classification enables the uncertainty to be accurately estimated via averaging over areas of high densities of the likelihood (Armero, Artacho, Lopez-Quilez, & Verdejo, 2011; Chipman, George, & McCullock, 1998; Denison, Holmes, Malick, &

* Corresponding author. Tel.: +44 07770918445.
E-mail address: Vitaly.Schetinin@beds.ac.uk (V. Schetinin).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.184

Smith, 2002; Duda, Hart, & Stork, 2000). The estimates of uncertainty are made over an ensemble of classification models obtained during Bayesian averaging. The use of Decision Trees (DTs) as classification models enables to select features which make the most significant contribution to the classification. The feature selection becomes important when prior information on EEG feature importance is absent or deficient. Besides, DTs are attractive classification models as experts can interpret them. In the case of ensembles, a single DT providing a Maximum Posterior can be selected for interpretation as we proposed in Schetinin et al. (2007).

The results of implementation of Bayesian averaging are critically dependent on the prior information and on the model parameter diversity in areas of averaging. When averaging is done over areas of interest with maximum likelihood, the resultant class posterior distribution is unbiased, and therefore the classification error is minimal. The use of prior information enables to specify the areas of interest and thus to improve diversity in model parameters.

Particularly, the prior information on EEG feature importance can be absent and so the areas of interest cannot be explicitly specified and then explored in detail (Domingos, 2000; Schetinin & Maple, 2007). Selection of EEG features has been shown improving the classification in Yom-Tov and Inbar (2002).

In our previous work (Jakaite & Schetinin, 2008), we attempted to mitigate the lack of prior information and proposed a new strategy for Bayesian averaging over DT models for predicting trauma survival. In this case of application, we observed that some screening tests (namely features) make a weak contribution to the model outcome and then we found that the DTs exploiting such weak tests can be discarded without affecting the accuracy of estimating the full class posterior distribution. In practice, it is important to reduce the number of features without an increase in the classification uncertainty, and the proposed method has been shown able to do achieve that.

The above findings motivated us to explore the discarding strategy in case of Bayesian assessment of newborn brain maturity from sleep EEG being represented by spectral power and statistical features. The importance of these features has not been explored yet in detail for a particular classification model such as DT. We will expect that the posterior information on EEG features will be effectively used within this strategy and the ensemble of DTs will be refined by discarding those models which exploit weak features. Similarly to the results obtained in our previous research, we will expect that the proposed strategy will reduce a portion of oversized DT models in the ensemble and the uncertainty in assessment will be decreased (Jakaite & Schetinin, 2008; Jakaite, Schetinin, & Maple, 2008).

Bayesian averaging over classification models is known as a theoretical methodology of achieving most accurate estimation of class posterior distribution. The estimate is calculated by integration of the posterior distribution over model parameters by using a stochastic integration known as Markov Chain Monte Carlo (MCMC) integration. The use of the Bayesian methodology allows experts to obtain the exhaustive information on uncertainty or risks in EEG assessment of newborn's brain. Therefore, the shape of the distribution becomes important for estimating the uncertainty in EEG assessment.

As part of this research, we will explore the shape of the class posterior distribution counted for a given PCA over DT models to answer the question whether a mismatch between the EEG estimate and PCA of the newborn causes a significant change in the shape. We assume that when PCA matches EEG estimate, the distribution shape tends to be symmetrical as the areas of interests are mainly located around one age category. On the contrary, for the mismatching cases the distribution becomes rather asymmetrical as the areas of interests are spread over different age categories. We will test our assumption on the EEG data to answer this question.

Overall, we expect to achieve the accuracy of the Bayesian assessment of brain maturity comparable to that obtained by experts. The accurate estimation of class posterior distribution provided by the Bayesian methodology will allow experts to obtain the exhaustive information on risk in EEG assessment of the newborn's brain maturity. The use of DT models which are transparent for users will allow EEG experts make new finding in the neurological assessment of newborn brain.

2. Problem statement

Typically, EEG experts assess the newborn brain maturity in terms of PCA measured in weeks. Most experts agreed that the physiological ages of newborns are known in the range ±2 weeks post conception. The weeks of PCA are most often counted on the base of information obtained from a questionnaire of the mother. Ultrasound dating has been shown more accurate than that and normally undertaken on the first and second triple-months. The dates are typically replaced by the ultrasound estimates if the difference exceeds ±1 week in the first triple and ±2 weeks, in the second triple (Hoffman et al., 2008).

The newborn EEGs are typically recorded via the standard C3T3-C4T4 electrodes during a few sleeping hours. In our case, the EEG recordings have been made by the polysomnograph Alice 3 with a sampling rate 100 Hz. The raw data have been then processed with the Fast Fourier Transform over each 6-s epochs to be represented by the standard spectral power bands: Subdelta (0–1.5 Hz), Delta (1.5–3.5 Hz), Theta (3.5–7.5), Alpha (7.5–13.5), Beta 1 (13.5–19.5 Hz), and Beta 2 (19.5–25 Hz).

For our experiments, the EEG features have been made consisting of two groups, basis and extension ones. The features of the basis group represent the relative and absolute values of the above six spectral power bands calculated for the two electrodes and their sum, making them 36. The features of the extension group represent the non-stationarity of an EEG recording estimated with our technique as shown in Jakaite, Schetinin, and Schult (2011). This technique estimates the distribution of the pseudo-stationary intervals in EEG. Using this technique, we made the extension group of features including the segment rate and 10 bins of the distribution histogram of the intervals ranging from 2-s to 20-s. These EEG features represent the information in the time domain. In particular, using the combined time and frequency EEG features has been shown improving the classification of EEG (Iscan, Dokur, & Demiralp, 2011). Finally, we added the ratio of slow-to-fast activities counted as the ratio of absolute spectral powers in Theta and Alpha bands, increasing the number of features in this group to 12.

The two feature groups together include 48 EEG features representing the EEG epochs. For our experiments, each EEG recording has been represented as a vector whose elements are the average values calculated over all epochs in the EEG recording. This is the typical way to represent each EEG recording as a vector in a multidimensional input space.

Note that although the above 48 features have been thought most informative for our experiments, we cannot state that there exists the prior information about the importance of either each feature or a feature combination considered within the given classification model. Therefore, using DT models for the Bayesian classification, we would explore the relative importance of the given EEG features and provide experts with the additional information about feature importance.

As mentioned in Section 1, in the absence of the prior information, the results of Bayesian classification will likely suffer from disproportionally sampling the posterior distribution, as we cannot expect that a multidimensional model space will be explored in detail, and the areas of interest will be properly explored within a reasonable time.