Robust Classification in RSVP Keyboard

Matt Higger, Murat Akcakaya, Umut Orhan, and Deniz Erdogmus

Northeastern University, 360 Huntington Ave, Boston, MA 02115 {Higger, Akcakaya, Orhan, Erdogmus}@ece.neu.edu

Abstract. To use in the Rapid Serial Visual Presentation (RSVP) KeyboardTM, a brain computer interface (BCI) typing system developed by our group, we propose a robust classification method of handling nonstationarity in the electroencephelography (EEG) data that is caused by artifacts and/or sensor failure. Considering the effect of these nonstationarities, we build a mixture data model to use as EEG evidence in the fusion with an n-gram language model to develop a robust classification algorithm. Using Monte Carlo simulations on the pre-recorded EEG data containing sections with or without intentionally generated artifacts we compare the typing performances of non-robust and robust classification methods in terms of speed and accuracy.

Keywords: BCI, ERP, Spelling.

1 Introduction

Locked-In Syndrome can isolate a person from those closest to them by taking away their ability to communicate. We focus on empowering those who are totally locked in, without control of any muscle group or eye gaze, by offering them a voice. Brain computer interfaces (BCIs) offer a promising avenue to do this. Generally, BCIs are methods which extract a person's intent through measurement of internal body signals. A common method, as we employ here, is to use the voltage of a person's scalp measured through Electroencephelography (EEG). EEG is a relatively cheap, portable, non-invasive way of measuring brain waves.

There are a number of EEG brain-phenomena which have been used to classify user intent. In motor imagery, a BCI system is designed to detect the signal generated by imaging the movement of a body part [1]. Additionally a steady state visually evoked potential (SSVEP) appears when a user is exposed to a periodic visual stimulus. Exposing a user to flickering checkerboard patterns, the induced SSVEP signals can be used to learn the user's gaze position from the frequency content of their brain waves [2].

Moreover, the EEG signals are sufficient for simple letter selection in the context of a typing algorithm for people with total-LIS. P300 signal, an event related potential (ERP) which occurs when a user is surprised by a circumstance, is commonly used for BCI spelling systems. P300 speller and Berlin BCI's Hexo Spell

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 443–449, 2013. © Springer-Verlag Berlin Heidelberg 2013

are well known examples of such systems [3], [4]. Different than these systems, in our approach, we utilize Rapid Serial Visual Presentation (RSVP), which presents the stimuli on the same location of the screen with temporal separation. The accuracy and speed of P300 typing systems suffer from low signal-to-noise ratio (SNR), the presence of artifacts in the signal and sensor failure and other effects that cause non-stationarity in the observed EEG signals. In this paper, we focus on a method to mitigate the influence of this non-stationarity on the typing performance.

The artifacts and/or sensor failure change the underlying distribution of the EEG data obtained from a BCI system causing a change in the optimal stimuli classification rule and degrading the system performance. Our goal is to develop a classification rule that is robust to changes in the assumed data distributions. To achieve this, we estimate the distribution of the data under different conditions, and using this distribution we develop our classification rule.

The rest of the paper is as follows. In Section 2, we explain the RSVP KeyboardTM, and then in Section 3, we develop the proposed robust classification rule. In Section 4, we demonstrate our experimental results, and conclude our discussion in Section 5.

2 RSVP KeyboardTM

The RSVP KeyboardTM consists of four main components: visual presentation, feature extraction, language modeling and the classifier used to select a symbol.

2.1 Visual Presentation

RSVP is a presentation technique in which visual stimuli are displayed as a temporal sequence at a fixed location on the screen. An example screen snapshot from the current RSVP Keyboard prototype is given in Figure 1.In the current study, RSVP contains random permutations of the 26 letters in English alphabet, a space symbol and a backspace symbol (a total of 28 symbols to choose from). We use the term "sequence" to mean a showing of all 28 symbols. If repetition is needed, all symbols are repeated multiple times to improve classification accuracy until a preset desired confidence level or a maximum number of repetition is reached. The process of repetition of sequences to choose a single symbol is named as an epoch. In an epoch, we make the assumption that the user shows positive intent for a single symbol.

2.2 Feature Extraction

The feature extraction starts by extracting stimulus-time-locked bandpass filtered EEG signals for each stimulus in the sequence. Since physiologically, the most relevant signal components are expected to occur within the first 500ms following the stimuli, the [0,500] ms portion of the EEG following each stimulus is extracted. At this stage it is important to design bandpass filters whose group



Fig. 1. RSVP Keyboard interface

delay does not shift the physiological response to outside this interval. A linear dimension reduction is applied on the temporal signals using Principal Component Analysis in order to to remove zero variance directions (i.e. zero-power bands based on the estimated covariance). The final feature vector to be classified is obtained as a concatenation of the PCA-projected temporal signals for each channel. Regularized Discriminant Analysis (RDA) [5] is used to further project the EEG evidence into scalar-feature for use in fusion with language model evidence.

RDA is a modification of quadratic discriminant analysis (QDA). QDA yields the optimal minimum-expected-risk Bayes classifier under the assumption of multivariate Gaussian class distributions. This classifier depends on the inverses of covariance matrices for each class, which are estimated from training data. To keep the calibration phase short few training samples are acquired - especially for the positive intent class. Therefore, the sample covariance estimates may become singular or ill-conditioned for high-dimensional feature vectors, which is the case here. RDA applies shrinkage and regularization on class covariance estimates. Shrinkage forces class covariances closer towards the overall data covariance as:

$$\hat{\Sigma}_C(\lambda) = \frac{(1-\lambda)\Sigma_C + \lambda\hat{\Sigma}}{(1-\lambda)N_C + \lambda\hat{N}} \tag{1}$$

Where λ is the regularization parameter, Σ_C, N_C are the class covariance estimate and number of samples for classes $C \in \{0, 1\}$ respectively. C = 0 is the non-p300 class. $\hat{\Sigma}, \hat{N}$ is the total covariance estimate and number of samples over all classes. Regularization is administered as:

$$\hat{\Sigma}_C(\lambda,\gamma) = (1-\gamma)\hat{\Sigma}_C(\lambda) + \frac{\gamma}{d}Tr[\hat{\Sigma}_C\lambda]I$$
(2)

where γ is the regularization parameter, /texttr[.] is the trace function and d is the dimension of the data vector.

After regularization and shrinkage, the covariance and mean estimates for each class are used in generating a scalar feature that minimizes expected risk under the Gaussianity assumption of class distributions. This is the log-likelihood ratio

$$\delta_{RDA}(x) = \log \frac{f_N(x;\hat{\mu}_1,\hat{\Sigma}_1(\lambda,\gamma)\hat{\pi}_1)}{f_N(x;\hat{\mu}_0,\hat{\Sigma}_0(\lambda,\gamma)\hat{\pi}_0)}$$
(3)

where $\mu_c, \hat{\pi}_c$ are estimates of class means and priors respectively; x is the data vector to be classified and $f_N(x; \mu, \Sigma)$ is the pdf of a multivariate Gaussian (normal) distribution.

2.3 Language Modeling

In letter-by-letter typing, we adopt an n-gram language models at the symbol level. These models estimate the conditional probability of a letter given by the n-1 previously typed letters. In this study, a 6-gram model that is trained using a one-million sentence (210M character) sample of the NY Times portion of the English Gigaword corpus. Corpus normalization and smoothing methods are described in [6]. Finally, we note that the backspace symbol is assumed to have a constant conditional probability of 0.05 and the conditional probabilities of the other symbols are normalized accordingly.

2.4 Classifier

Using the class conditional score and the language model probabilities in a naive Bayes' rule based fusion model, we compute the posterior probabilities of symbols given all the evidence. We compute these probabilities for each symbol after every sequence, and a decision is made if one symbol probability reaches a desired confidence level or number of repetitions exceeds a predefined limit.

3 Robust Classifier

In the classifier, the class conditional score distributions are used assuming that these distributions remain stationary during a typing session. However, possible changes in the distribution of the EEG data, possibly due to artifacts or sensor failure, should be incorporated in the score distribution. For example, as we also explain in Section 4, we apply our method on artifact reduction assuming artifacts as possible reasons for changes in the distribution. We introduce a variable a which describes the artifact class of a particular trial. Artifact classes include a control group (no artifacts present), eye blink, jaw movement and smiling. For use in the language model fusion, we compute the score conditional distributions for the mixed conditional score distribution as

$$P(\delta_{RDA}(x)|c) = \Sigma_i P(\delta_{RDA}(x)|c, a_i) P(a_i)$$
(4)

where, *i* is the artifact index, c = 0 or 1 is the class label, $P(a_i)$ is the prior for artifact a_i . For each class and artifact $P(\delta_{RDA}(x)|c, a_i)$ is computed using (2.2).

4 Experiments

Four healthy operators participated in this study. For each subject, four RSVP sessions with pre-designated targets were performed using a 16-channel g.USBamp and g.Butterfly electrodes (g.Tec, Graz, Austria) in one sitting. The second session was the control session, while the first, third and fourth sessions had the subjects produce intentional jaw movement, eye blinks, and face muscle artifacts, respectively. Subjects continued to attend to the RSVP presentation during all sessions. This data is used to build and test robust and non-robust fusion models using 10-fold cross validation as explained in Section 3

We perform Monte Carlo simulations on multiple pre-recorded calibration data sets to build kernel density estimates (KDEs) of the RDA score distribution for target symbol present and not-present conditions.

We select ten different sentences and aim to spell a phrase in each sentence (called the copy phrase task). Task difficulty is determined by requiring each letter of the target phrase to have a likelihood ratio against the highest likelihood competing non-target letter within a specified interval: (1) Hard: (0.3, 0.5], (2) Very hard: (0, 0.3].

In summary, we model typing performance by building a distribution of RDA scores from real training data under different artifact conditions. This model is then simulated typing 10 sentences 15 times to compare the performances of robust and non-robust classifiers. We report our results in terms of typing accuracy and duration (total seconds per word completion), see Figures 2 and 3. For reference, we include the area under the curve (AUC) values for each subject under all artifact conditions in 1.

| | Subject 1 | Subject 2 | Subject 3 | Subject 4 |
|--------------|-----------|-----------|-----------|-----------|
| No-Artifact | .7644 | .8298 | .6488 | .8103 |
| Jaw Movement | .6079 | .8026 | .6370 | .6527 |
| Smile | .7105 | .8423 | .6506 | .7023 |
| Eye Blink | .6561 | .7641 | .4710 | .7373 |

Table 1. AUC values

4.1 Typing Accuracy

As can be noted in Figure 2, typing accuracy changes dramatically between subjects. In the simulation, as with other trials we've performed, subject 3 struggles to produce accurate classifications. Additionally, we note that robust classification consistently outperforms non-robust methods. The performance advantage of our method is correlated to the magnitude of the difference in AUC between the control and artifact classes. In other words, the stronger the drop in AUC when an artifact is introduced (Table 1), the greater the performance benefit of using robust fusion.



Fig. 2. Accuracy vs artifact type

4.2 Typing Duration

From Figure 3, we immediately notice that the robust case typically types faster than the non-robust case. Additionally, considering the AUC values from Table 1 and the results from Figure 3, we notice that higher AUC values offer quicker typing performance. Both these effects share a common motivation. The typing system repeats sequences until a sufficiently high confidence threshold is reached. Accurate typing, because of robust methods or high user AUC, will yield fast typing.



Fig. 3. Typing duration vs artifact type

5 Conclusions

We designed a robust classification method for ERP detection in a BCI typing paradigm. We tested the proposed method on the RSVP KeyboardTM, which is an in-house BCI typing system. Considering the possible changes in the EEG data, we developed a mixture density model for class conditional EEG evidence to use in the fusion with n-gram language model. To compare the robust and non-robust classification methods, using pre-recorded calibration data, we simulated the performance of four subjects typing 10 sentences 15 times and reported results on accuracy and speed of their typing.

Each of our simulations was run under a single artifact class (rather than a mixture of multiple classes). When implemented with a true mixture of artifact classes we observed nearly identical results between robust and non-robust methods. We suggest this is due to the ability of our classifier to accumulate additional EEG evidence when an input doesn't reach the confidence threshold. In the true mixture case, where artifacts aren't very frequent, the classifier is bound to receive useful information during the following sequences. For further analysis, we are interested in examining the cause of mis-classified symbols. We hypothesize that the risk in artifacts is not readily seen in their prior distribution as artifacts frequently occur in bursts during operation. Future artifact class models which are conditioned on previous artifact classes, allowing for bursts of artifacts to occur while still keeping artifact priors at reasonable levels, will be studied.

References

- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.R.: Optimizing spatial filters for robust eeg single-trial analysis. IEEE Signal Processing Magazine 25(1), 41–56 (2008)
- Nezamfar, H., Orhan, U., Erdogmus, D., Hild, K., Purwar, S., Oken, B., Fried-Oken, M.: On visually evoked potentials in eeg induced by multiple pseudorandom binary sequences for brain computer interface design. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2044–2047 (2011)
- Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: Toward enhanced P300 speller performance. Journal of Neuroscience Methods 167(1), 15–21 (2008)
- 4. Treder, M., Blankertz, B. (C)overt attention and visual speller design in an erpbased brain-computer interface. Behavioral and Brain Functions 6(1), 28 (2010)
- Friedman, J.H.: Regularized discriminant analysis. Journal of the American Statistical Association 84(405), 165–175 (1989)
- Roark, B., de Villiers, J., Gibbons, C., Fried-Oken, M.: Scanning methods and language modeling for binary switch typing. In: The NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies, p. 2836. Association for Computational Linguistics (2010)