A NEW MULTIPLE-KERNEL-LEARNING WEIGHTING METHOD FOR LOCALIZING HUMAN BRAIN MAGNETIC ACTIVITY

T. Takiguchi[⋆†], *T. Imada*[†‡], *R. Takashima*[⋆], *Y. Ariki*[⋆], *J.-F. L. Lin*[†], *P. K. Kuhl*[†], *M. Kawakatsu*[‡], *M. Kotani*[‡]

[⋆]Graduate School of System Informatics, Kobe University, Japan
[†]Institute for Learning and Brain Sciences, University of Washington, USA
[‡]Tokyo Denki University, Japan

ABSTRACT

This paper shows that pattern classification based on machine learning is a powerful tool to analyze human brain activity data obtained by magnetoencephalography (MEG). We propose a new weighting method using a multiple kernel learning (MKL) algorithm to localize the brain area contributing to the accurate vowel discrimination. Our MKL simultaneously estimates both the classification boundary and the weight of each MEG sensor; MEG amplitude obtained from each pair of sensors is an element of the feature vector. The estimated weight indicates how the corresponding sensor is useful for classifying the MEG response patterns. Our results show both the large-weight MEG sensors mainly in a language area of the brain and the high classification accuracy (73.0%) in the $100 \sim 200$ ms latency range.

***Index Terms*—** brain area, magnetoencephalography, brain activity, weighting, kernel learning

1. INTRODUCTION

Non-invasive measurements using magnetoencephalography (MEG) have recently been used to study how stimulus features are processed in the human brain. In particular, because neural electric activity of the brain associated with speech and language stimuli happens in a time frame of milliseconds, the high temporal resolution of MEG is required for measuring rapid changes in brain activity during speech perception. Research carried out with MEG has reported left hemisphere dominance for processing of vowels in right-handed subjects [1], and the prominent N1m wave of the auditory-evoked field has been shown to exhibit sensitivity to a variety of acoustic attributes of the speech signal [2], as well.

Recently, application of pattern recognition methods to neuromagnetic responses has created much interest, and progress has been made through the use of machine learning, such as support vector machines (SVMs) [3, 4, 5]. SVMs are efficient tools for automatic recognition, but neuroscience research requires not only classification tools (that have high accuracy) but also analysis tools that can locate both the dominant area of the brain, showing strong activity related to speech and language, and the significant time frame, exhibiting this increased brain activity.

A multiple kernel learning (MKL) algorithm is a machine-learning-based technique for learning proper weights of the corresponding kernels, in using multiple classifiers with a kernel, such as SVM. MKL has been used as an integration method by calculating appropriate weights corresponding to each kernel, while classical kernel-based methods (such as SVMs) are based on a single kernel only. In image processing research field, object recognition methods based on MKL have been proposed for integrating image features [6].

In this paper, we present a new weighting method for the MKL algorithm, where the weight is associated with each MEG sensor. In our approach, MKL was applied to MEG responses or amplitudes, to localize brain areas that contribute to the accurate decoding of vowels. Sixty-one MEG amplitudes, each calculated from each of 61 pairs of MEG sensors (in total 122 MEG sensors), constituting a 61-dimensional feature vector, are separately weighted; each weight value calculated by MKL indicates how useful each MEG sensor pair is for classifying the MEG responses to vowel recognition. To identify the MEG sensors or brain areas important for vowel recognition in a subject-independent (subject-open) fashion, the weights were averaged across subjects.

2. RECORDING OF MEG RESPONSES TO VOWELS

Eight right-handed volunteers (4 males and 4 females; 21-25 years old) were recruited as subjects after obtaining consent forms from them. All were native Japanese speakers with normal hearing.

We used two speech sounds (Japanese vowels), /a/ and /o/, to explore subject's vowel recognition process in the brain. These 200-ms auditory stimuli were delivered to the subject's right ear through a plastic tube with a random interstimulus interval between 1,300 and 1,500 ms. The subject's task was to press a reaction key with the index finger when the subject identified the stimulus /a/ and another reaction key with the middle finger when the subject identified the stimulus /o/.

Neuromagnetic data were recorded by a 122-channel whole-scalp Neuromag MEG system in a magnetically shielded room. The MEG signal was sampled at 497 Hz for 1,200 ms including a 100-ms pre-stimulus baseline; more than 80 epochs were averaged to increase the S/N ratio. A low-pass filter with a cutoff frequency of 40 Hz was used in calculating the feature vector. Epochs in which the magnetic signal exceeded an absolute amplitude variation of 3,000 fT/cm were discarded. Eye-movement artifacts were also automatically removed (threshold = 150 μV).

Feature extraction was applied to a 996-ms MEG signal. Since the mean reaction times, however, for /a/ and /o/ were 495.1 ms (SD = 51.7) and 497.3 ms (SD = 46.8), respectively. The MEG feature vectors up to 450 ms were used to analyze the MEG response pattern to localize the brain activation during recognizing vowels.

3. FEATURE EXTRACTION

The signal obtained by averaging over 80 MEG epochs was converted (using a feature extraction transformation) into a representation more amenable to subject-independent recognition. As inter-subject variability in MEG signals degrades the recognition accuracy of a machine learning system, MEG magnitude was normalized by the following statistical method.

The MEG signal at time t is represented by

$$\mathbf{x}(t) = [x_1(t), \cdots, x_m(t), \cdots, x_M(t)]^T \quad (1)$$

where $x_m(t)$ denotes the observation at the m-th sensor, and the symbol M denotes the total number of MEG sensors. To avoid canceling problems due to the polarity difference between subjects, the MEG magnitude was first calculated by the following Eq. (2), which is a vector magnitude of paired vertical and horizontal sensors.

$$y_j(t) = \sqrt{x_i^2(t) + x_{i+1}^2(t)} \quad (2)$$

where $y_j(t)$ ($1 \leq j \leq M/2$) is the magnitude feature.

To reduce the inter-subject variability problems in MEG magnitudes, the magnitude feature is normalized to have zero mean and unit variance.

$$\hat{y}_j(t) = (y_j(t) - \bar{y}_j)/\sigma_j \quad (3)$$

$$\bar{y}_j = \frac{1}{T}\sum_t y_j(t), \ \sigma_j = \sqrt{\tfrac{1}{T}\sum_t (y_j(t) - \bar{y}_j)^2} \quad (4)$$

where $\bar{y}_j$ denotes the mean magnitude feature, T denotes the total number of samples for each averaged epoch, and σ_j denotes the standard deviation. Figure 1 shows average MEG response magnitudes from a sensor over the left language area of a typical subject. The deflection at 100 ms is clearly strong for both stimuli /a/ and /o/, but the difference between /a/ and /o/ is also seen between 150 and 250 ms.

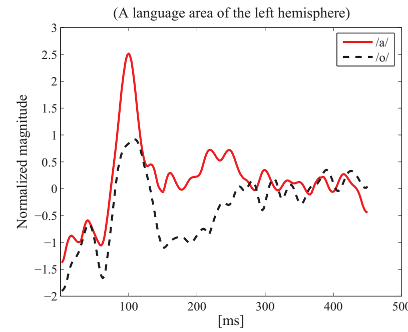

Fig. 1. Normalized MEG-magnitude features obtained at a single MEG-sensor site with a pair of MEG sensors over the left language area of a typical subject.

The normalized MEG magnitude feature at each MEG sensor, obtained from Eq. (3), constituted 61-dimensional MEG-magnitude feature vector, as shown in Eq. (5), for further analysis or classification using a multiple kernel learning algorithm.

$$\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \cdots, \hat{y}_{M'}(t)]^T, \ M' = M/2 \quad (5)$$

4. MEG-SENSOR WEIGHTING AND CLASSIFICATION BASED ON MKL

The MKL algorithm has been used to integrate multiple conventional kernel-based methods, such as SVMs, which rely only on a single kernel (See Fig. 2 upper panel) by assigning appropriate weights to those multiple component kernels. Our MKL approach was developed to localize brain areas associated with the subject's task, namely the accurate decoding of vowels, by assigning independent weights to each MEG sensor, where the larger the MEG-sensor weight is, the more important role the brain activity underneath the MEG-sensor plays.

In an MKL framework, a total or global kernel function is defined as a linear combination of the base kernels.

$$k(\hat{\mathbf{y}}(p), \hat{\mathbf{y}}(q)) = \textstyle\sum_l \beta_l k_l(\hat{\mathbf{y}}(p), \hat{\mathbf{y}}(q)) \quad (6)$$

Here k_l is the l-th base kernel computed from the p-th and q-th samples of the feature vector $\hat{\mathbf{y}}(p)$ and $\hat{\mathbf{y}}(q)$, and the non-negative coefficient β_l represents the weight of the base kernel. The MKL approach for SVMs has been originally used to improve the classifier performance by combining various classifiers with different kernels, each receiving the same feature vector. In recent image recognition research, however, the MKL approach started being used for the purpose of feature vector selection or weighting. For this purpose, the weight is independently trained for each base kernel receiving different feature vector [6]; see Fig. 2 lower panel. Since the weight is different depending on the feature vector, call

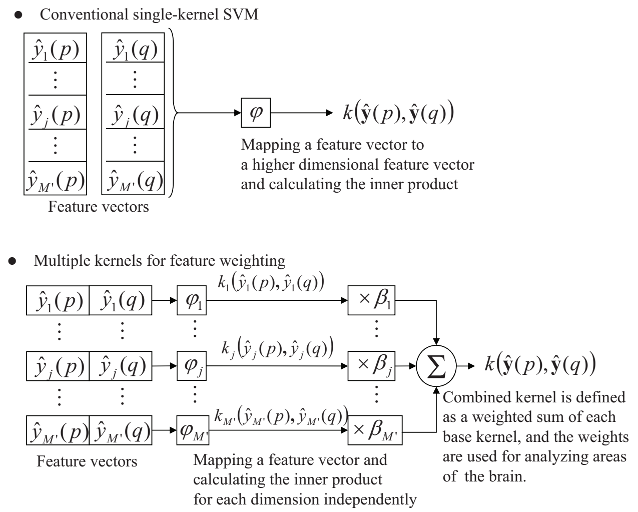

Fig. 2. A conventional SVM and a new weighting method based on MKL.

this method a feature-weighting MKL method in this paper. Simplifying this new approach, we have previously proposed a method for single-channel sound source localization using the acoustic transfer function, where each dimension of the acoustic transfer function is selected using [7].

In this paper, we employ a simplified version of the above mentioned feature-weighting method, similar to that proposed in [7]. In this simplified version, each dimension or element of the MEG feature vector corresponds to an MEG magnitude at a pair of sensors, which in turn associated with a brain area underneath this pair of MEG sensors; the weight assigned to each dimension is trained using MKL; and each weight also corresponds to each base kernel (see Fig. 2 lower panel). Although SVM can work as an efficient tool for classifying multiple MEG activities, due to its single kernel nature, it is difficult for a single kernel SVM to localize the brain areas, which MEG data would be able to provide. On the other hand, the MKL-SVM introduced in this paper can localize the brain areas that is involved in discriminating two vowels, by estimating the weight of each dimension of the feature vector because the large weight indicates that there is information useful for classifying the MEG responses. In this paper, the feature weights are trained by defining the base kernel for each of 61 dimensions of the feature vector as follows:

$$k(\hat{\mathbf{y}}(p), \hat{\mathbf{y}}(q)) = \textstyle\sum_j \beta_j k_j(\hat{y}_j(p), \hat{y}_j(q)) \tag{7}$$

The kernel weight β_j is trained on an SVM framework (i.e., maximum-margin-based scheme). In the SVM framework, the MKL criterion is defined by the following objective func-

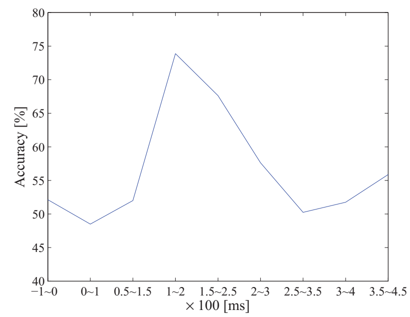

Fig. 3. Classification accuracy using MKL in each 100-ms latency range.

tion.

$$\begin{aligned} \max_{\alpha,\beta} \quad & \textstyle\sum_p \alpha(p) - \frac{1}{2}\sum_{p,q} \alpha(p)\alpha(q)z(p)z(q) \\ & \qquad \cdot \textstyle\sum_j \beta_j k_j(\hat{y}_j(p), \hat{y}_j(q)) \\ s.t. \quad & \begin{cases} \sum_p z(p)\alpha(p) = 0, & 0 \le \alpha(p) \le C \\ \sum_j \beta_j = 1, & \beta_j \ge 0 \end{cases} \end{aligned} \tag{8}$$

Here $\alpha(p)$ is the Lagrange coefficient, and $z(p) = \{+1, -1\}$ denotes the class label of example $\hat{\mathbf{y}}(p)$. C determines the trade-off between the margin and training data error. In Eq. (8), both $\alpha(p)$ and β_j are optimized by a two-step iterative procedure. In the first step, β_j is fixed, and $\alpha(p)$ is updated by a standard SVM solver. In the second step, $\alpha(p)$ is fixed, and β_j is optimized by a projected-gradient scheme. In this way, the feature weights and the classification boundary are trained simultaneously.

5. SUBJECT-INDEPENDENT ANALYSIS ON RECORDED MEG DATA

5.1. Analysis conditions

The MKL-based analysis was evaluated on neuromagnetic responses to recognition of the vowel sounds /a/ and /o/. The total number of subjects was 8. We used leave-one-subject-out cross-validation, where the number of training subjects was 7, and the remaining one subject was used as the test data. We then repeated this until all the subjects were tested. The MKL algorithm was independently applied to every latency range, where a single latency range contains 50 samples (about 100 ms with a sampling frequency of 497 Hz). The frame period was set to 25 samples, meaning that the 100-ms latency range moves every about 50 ms from 0 ms until 350 ms. Since the reaction times for both speech sounds were about 500 ms, we assumed the discrimination was finished by 400 ms at latest, resulting in the final latency range from 350ms to 450 ms for further analysis. A Gaussian kernel was employed as the kernel function, and the hyper parameter C in Eq. (8) was 1.

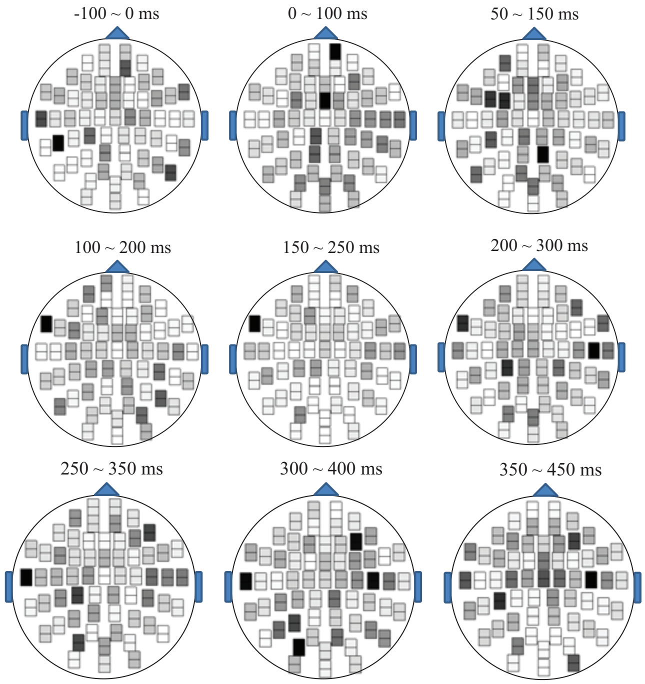

Fig. 4. MEG-sensor weighting based on MKL. There are 9 top-view circle heads with nose upward.

5.2. Analysis results

Figure 3 shows the average classification accuracies obtained from subject-independent (subject-open) analysis. As can be seen in this figure, the classification accuracy first increased as a function of time, reached a maximum vale of 73.9% in the latency range between 100 and 200 ms, and then decreased to a chance level (50%) in the latency range between 250 and 350 ms. The classification accuracy of 73.9% gives 3.5% higher than the best accuracy of a single kernel SVM. Welch's unpaired t-test (at the 2.5% significance level) shows that the MKL method gives significantly better classification performance in the latency range between both 100 and 200 ms, and 150 and 250 ms compared with that at the 100-ms pre-stimulus baseline.

To localize MEG sensor important for MEG activity pattern classification using MKL, which were considered to have contributed to the processing of vowel recognition, the MEG sensor weights (β_j in Eq. (7)) obtained from the MKL method are displayed on a topological plot of the scalp in Fig. 4. Figure 4 shows color-coded average weights for each MEG sensor in each latency range; more important or more highly weighted MEG sensors for classifying neuromagnetic responses are shown in darker colors; the black areas indicate that this area of the brain played an important role in classification of neuromagnetic responses to vowel recognition. The larger weights in the latency range both between 100 and 200 ms and between 150 and 250 ms, where high accuracy was achieved, are seen to be in the left language area.

6. CONCLUSION

We presented a new MEG-sensor weighting method using a multiple kernel learning algorithm for analyzing areas of the brain that contributed to the accurate decoding of two vowels. Our subject-independent (subject-open) analysis results showed a high classification accuracy of 73.9% obtained in the latency range between 100 and 200 ms, where we observed strong MEG waveform peaks, for a two-vowel recognition task. The classification accuracy of 73.9% obtained by our MKL method was 3.5% higher than the best accuracy of a conventional single kernel SVM. The brain area covered by the MEG sensors with the larger weight obtained by our MKL method corresponded to the language area of the left hemisphere. As the magnetic fields generated by brain activity are extremely weak and usually largely contaminated by external magnetic noises, we will have to develop a noise-robust feature extraction method.

7. ACKNOWLEDGMENTS

This research was funded by Japan Society for the Promotion of Science and National Science Foundation Science of Learning Center grant to the University of Washington's LIFE Center (SBE-0354453 to P.K. Kuhl).

8. REFERENCES

[1] L. Gootjes, T. Raij, R. Salmelin, and R. Hari, "Left-hemisphere dominance for processing of vowels: a whole-scalp neuromagnetic study," *NeuroReport*, vol. 10, pp. 2987–2991, 1999.

[2] I. Miettinen, H. Tiitinen, P. Alku, and P. J.C. May, "Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds," in *BMC Neuroscience*, 2010, pp. 11–24.

[3] N. F. Ince, F. Goksu, G. Pellizzer, A. Tewfik, and M. Stephane, "Selection of spectro-temporal patterns in multichannel MEG with support vector machines for schizophrenia classification," in *EMBS*, 2008, pp. 3554–3557.

[4] F. Asano, M. Kimura, T. Sekiguchi, and Y. Kamitani, "Classification of movement-related single-trial MEG data using adaptive spatial filter," in *ICASSP*, 2009, pp. 357–360.

[5] J. M. Baker, A. M. Chan, K. Marinkovic, E. Halgren, and S. S. Cash, "Robust representations of cortical speech and language information," in *ICASSP*, 2011, pp. 785–788.

[6] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *ICCV*, 2007, pp. 1–8.

[7] R. Takashima, T. Takiguchi, and Y. Ariki, "Feature selection based on multiple kernel learning for single-channel sound source localization using the acoustic transfer function," in *ICASSP*, 2011, pp. 2696–2699.