Model Individualization for Real-Time Operator Functional State Assessment

Guangfan Zhang
Intelligent Automation, Inc.

Roger Xu
Intelligent Automation, Inc.

Wei Wang
Intelligent Automation, Inc.

Aaron A. Pepe
Intelligent Automation, Inc.

Feng Li
Old Dominion University

See next page for additional authors

Original Publication Citation

Zhang, G., Xu, R., Wang, W., Pepe, A., Li, F., Li, J., McKenzie, F., Schnell, T., Anderson, N., & Heitkamp, D. (2012). Model individualization for real-time operator functional state assessment. In S. J. Landry (Ed.), *Advances in Human Aspects of Aviation* (pp. 417-426). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/b12321-48/model-individualized-real-time-operator-functional-state-assessment

Model Individualization for Real-Time Operator Functional State Assessment

Guangfan Zhang, Roger Xu, Wei Wang, and Aaron A. Pepe

Intelligent Automation, Inc.
{gzhang, hgxu, wwang, apepe}@i-a-i.com

Feng Li, Jiang Li, and Frederick McKenzie

Old Dominion University
jli@odu.edu; feng.odu@gmail.com; rdmckenz@odu.edu;

Tom Schnell, Nick Anderson, Dean Heitkamp

University of Iowa
thomas-schnell@uiowa.edu; nick@andersonaviation.net; dean-heitkamp@uiowa.edu

ABSTRACT

Proper assessment of Operator Functional State (OFS) and appropriate workload modulation offer the potential to improve mission effectiveness and aviation safety in both overload and underload conditions. Although a wide range of research has been devoted to building OFS assessment models, most of the models are based on group statistics and little or no research has been directed towards model individualization, i.e., tuning the group statistics based model for individual pilots. Moreover, little emphasis has been placed on monitoring whether the pilot is disengaged during low workload conditions. The primary focus of this research is to provide a real-time engagement assessment technique considering individual variations in an aviation environment. This technique is based on an advanced machine learning technique, called enhanced committee machine. We have investigated two different model individualization approaches: similarity-based and dynamic ensemble selection-based. The basic idea of the similarity-based technique is to find similar subjects from the training data pool and use their data together with the limited training data from the test subject to build an individualized OFS assessment model. The dynamic ensemble selection dynamically select data points in a validation dataset (with labels) that are adjacent to each test sample, and evaluate all the trained models using the identified data points. The best performing models will be selected and maximum voting can be applied to perform individualized assessment for the test sample. To evaluate the developed approaches, we have collected data from a high fidelity Boeing 737 simulator. The results show that the performance of the dynamic ensemble selection approach is comparable to that achieved from an individual model (assuming sufficient data is available from each individual).

Keywords: Operator Functional State, Engagement, EEG, Machine Learning, model individualization

1 INTRODUCTION

Research on Operator Functional State (OFS) assessment has attracted considerable attention in the recent decade. According to the North Atlantic Treaty Organisation (NATO), OFS is defined as the multidimensional pattern of human psychophysiological condition that mediates performance in relation to physiological and psychological costs (Hockey 2003). In commercial flights (especially long-haul flights), pilots often experience short periods of high workload during pre-flight preparations, takeoff and landing, and long periods of very low workload as the aircraft cruises enroute toward the destination with autopilot. For high workload tasks, an operator has to deal with large amounts of information, make multiple decisions, or carry out critical actions within a short period of time. As a result, the OFS of the operator may not meet the task demand and human errors may lead to disastrous consequences. On the other hand, in commercial flights (especially long-haul flights), pilots often experience long periods of low workload as the aircraft cruises enroute toward the destination with the aircraft on autopilot. Pilots can easily become disengaged as they may be less attentive under low workload. The disengaged pilots may not be able to properly handle the unexpected events, such as turbulence, equipment failure/malfunction and even potential collisions with other aircraft. Thus, proper assessment of OFS and workload modulation offer the potential to improve mission effectiveness and aviation safety in both overload and underload conditions (Schmorrow & Kruse, 2002; Schmorrow et al., 2005).

In the past decade, a wide range of research has been undertaken regarding OFS assessment. Many existing studies utilized psychophysiological measurements to index the level of cognitive demand associated with a task (Boucsein & Backs, 2000), fatigue (Smith et al., 1999 & Trejo et al., 2005), engagement (Stevens, 2007 & Pope et al., 1995) and other functional state dimensions (Hockey 2003). However, model individualization has not been well addressed. It is often difficult, if not impossible; to train an OFS assessment model for each individual pilot due to lack of sufficient training data. For each individual, there are individual differences, commonly referred to idiosyncratic regularities of the physiological reaction or Individual Response Specificity (IRS) (Engel, 1960; Marwitz & Stemmler, 1998). The consistency of the individual response specificity over time was also documented by Forster (Foerster, 1985).

However, current methods pay little attention to individual variations. A generalized OFS assessment model is usually built based on the data from a large amount of training subjects and is then applied directly (or with a minimum adaptation) to a new subject of interest (test subject). However, due to large individual variations, the generalized model often yields poor performance. Olofsen et al. (2010) categorized two model individualization techniques. The first involves basic research, which aims to discover the individual differences of physiological responses reflecting the OFS. Based on the knowledge and understanding of the physiological response, model parameters can be adjusted accordingly to address individual variations. The second approach doesn't rely on the understanding of the physiological differences and only involves statistical analysis/modeling approaches. Currently, the understanding of the nature of OFS is limited and many fundamental issues, including individual variations in physiological response, are not well understood. Thus, most of the current research focuses on the statistical approaches. Rajaraman et al. (2009) has developed a method for developing individualized biomathematical models for predicting cognitive performance impairment of individuals subjected to total sleep loss. Their method systematically customizes the parameters in the two-process model of sleep regulation for an individual by optimally combining the performance information with a priori performance information using a Bayesian framework. Zhang et al. (2009) presented a similarity-based approach for model individualization. This approach finds similar subjects from the training data pool and uses their data together with the limited data from the test subject to build an individualized OFS assessment model. The idea is based on the assumption that if the test subject's data is similar to the data of a training subject in one or more specific conditions, he or she will have similar behaviors to the training subject in other conditions. Thus all the training data from this training subject can be used to build a model for the test subject.

This paper presents a framework for real-time individualized OFS (engagement, more specifically) monitoring in an aviation environment. Two major components are included. The first component is a base OFS assessment component, which is based on an advanced machine learning technique, enhanced committee machine. A committee machine is a strategy to improve classification/regression performance by combining responses from multiple committee members. The enhanced committee machine integrates a bootstrapping technique, an advanced feature selection method and a Neural Networks-based classification method to build base classifiers/committee members (Zhang 2009). By aggregating outputs from the committee members, the final engagement decision can be more robust and accurate. The second component is the model individualization component, which performs individualized engagement assessment. We further explore the similarity-based individualization technique by comparing different similarity measures and investigate a novel dynamic ensemble selection-based approach for model individualization.

The dynamic ensemble selection framework dynamically selects a subset of committee members/classifiers from tens/hundreds (or more) of committee members. The selection is based on the performance of the committee members on the limited validation data (with labels). The rational of this design is based on the assumption that if the test sample is similar to its local validation samples in the feature space and we select a subset of trained committee/members classifiers who perform well on the local validation samples, we may achieve a good engagement assessment results for the test data point. If there is limited data from the test subject with labels (known engagement states), it can also be included in the validation dataset to evaluate the performance of each classifier. The first few top classifiers are then selected to classify the test sample by majority voting (Giacinto, 2001).

To evaluate the developed approach, we have collected data from a high fidelity Boeing 737 simulator, including flight technical data, psycho/physiological signals and behavioral measures. An extensive feature study has been performed to extract promising and robust features. The results show accurate engagement/disengagement detections, making it suitable for real-time assessment. Both of the model individualization techniques have been evaluated and the results show that the performance of the dynamic ensemble selection approach is comparable to that achieved using an individual model assuming sufficient data is available from each individual for training. We have also shown that the similarity-based technique has the potential to individualize the OFS assessment model. However, the performance achieved based on the limited subjects doesn't show an improvement comparing to the results achieved from a generalized model. It may be due to the limited number of subjects available for performance evaluation. We will further investigate this method in the future.

The remainder of the paper is organized as follows. Section 2 describes model individualization based on enhanced committee machine for real-time engagement assessment. Section 3 describes the flight simulation configuration and experiment design. Section 4 shows performance evaluation results. Section 5 concludes the paper.

2 MODEL INDIVIDUALIZATION BASED ON ENHANCED COMMMITTEE MACHINE

In this research, based on the enhanced committee machine framework (Zhang et al. 2009), we have developed two model individualization techniques: similarity-based and dynamic ensemble selection-based.

2.1 Enhanced Committee Machine

A committee machine is a strategy to improve classification/regression performance by combining responses from multiple committee members. Different algorithms can function as committee members, for example, Neural Networks (NNs), Gaussian Mixture Models (GMMs), and Support Vector Machines (SVMs). If committee members have the diversity property, i.e. they are unlikely to make errors in the same feature space, the errors from individual committee members will be canceled by each other to some extent. Furthermore, since the committee machine "averages" its individual member's estimation, the variance of the committee machine can be significantly reduced. As a consequent, the performance of the combination of the estimation from each committee member is often superior to that of its committee members (Zhang et al., 2009).

In order to obtain the diversity property, committee members are usually trained by bagging or boosting (Breiman, 1996) if training data size is fixed, or by independent datasets if they are available. We enhanced the committee machine method by combining it with an advanced feature selection algorithm to select different features for each committee member. Each of the members is trained on the features selected by the Piecewise Linear Orthogonal Floating Search (PLOFS) feature selection algorithm, which is a wrapper type of algorithm and is as fast as a filter approach. It consists of a piecewise linear network (PLN) and the floating search procedure through the orthogonal space (Li et al., 2006).

In this paper, we implemented the committee machine using a Multi-Layer Perceptron (MLP) neural network trained by the standard Back Propagation (BP) algorithm as the base classification model. There exist many ways to combine the outputs of committee members. If only the label output is available for each member, a majority vote scheme is often used. It is also worth noting that we delete some of the committee members with a high bias, i.e., having low training accuracies, to enhance the OFS assessment performance.

2.2 Model Invidualization

It is known that physiological signals from different individuals usually have different characteristics. Therefore, a generalized model built on data from other individuals may not perform well. When an individual's data is not sufficient to build an individual's model, an alternative option is to individualize the generalized model for each individual. We have investigated two approaches for model individualization in this research: similarity-based and dynamic ensemble selection.

The architecture of the two model individualization techniques based on enhanced committee machine for engagement assessment is shown in Figure 1. It is common that both methods perform model individualization by selecting a subset of committee members, from similar subjects (similarity based) or best performing committee members based on performance evaluation results using a validation dataset (dynamic ensemble selection). We will introduce both methods in this section.

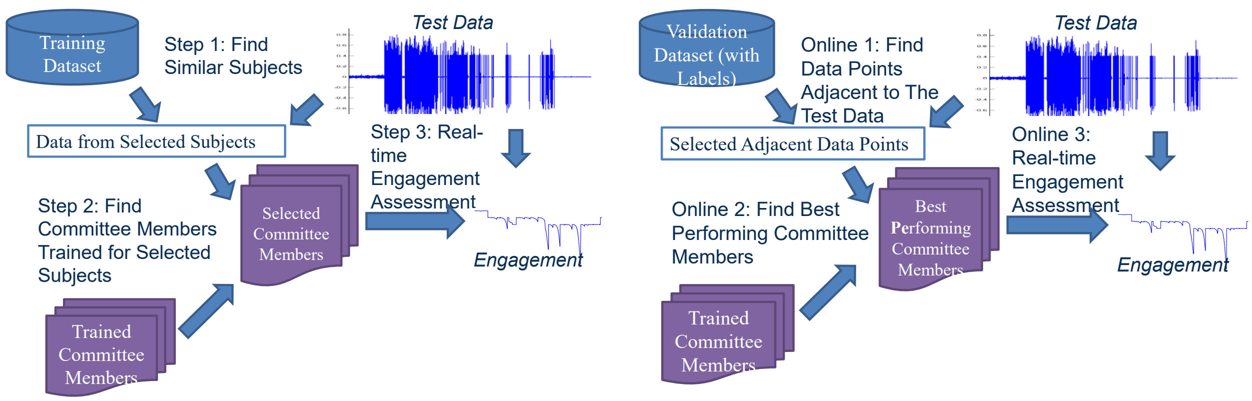

Figure 1 Model individualization based on enhanced committee machine: (a) similarity-based (left) and (b) dynamic ensemble selection (right).

2.2.1 Similarity-based Model Individualization

In many cases, training data for an individual is limited and is expensive to collect. It may be infeasible to train an individualized OFS monitoring model using only the individual training data. Therefore, we propose an approach to build an individualized OFS monitoring model by identifying training subjects that have similar physiological responses and extracting their data for model individualization.

Ideally, for each individual being tested, we assume that we can find one or more "similar" subjects among all the subjects with training data. In each functional state that the individual has experienced, there may exist one or more subjects that have similar training data in the same functional state. The purpose of computing subject distance is to find such subject(s) in a specified functional state. After scanning all the functional states of the individual, we can select a subset of subjects that are similar to the individual in some functional states based on the similarity metrics. All the trained committee members from these subjects can be extracted to form an individualized committee machine.

The similarity can be measured by different metrics. In this paper, we investigated five different similarity measures implemented in Matlab: t-test, entropy, Bhattacharyya distance, ROC and Wilcoxon test (MATLAB 7.11, The MathWorks Inc., Natick, MA, 2010).

- t-test: Absolute value two-sample t-test with pooled variance estimate.
- Entropy: Relative entropy, also known as Kullback-Leibler distance or divergence.
- Bhattacharyya: Minimum attainable classification error or Chernoff bound.
- ROC: Area between the empirical receiver operating characteristic (ROC) curve and the random classifier slope.
- Wilcoxon: Absolute value of the u-statistic of a two-sample unpaired Wilcoxon test, also known as Mann-

Whitney.

2.2.2 Dynamic Ensemble Selection

To perform dynamic ensemble selection for model individualization, we first train a set of committee members using training data. Next, a validation dataset is selected, such as limited data from the test subject and a subset of data from other subjects. For each test sample, its nearest neighbors are selected dynamically in the validation dataset and all the trained committee members are evaluated using the identified data points. The best performing committee members are selected to form a committee machine for the test sample (Figure 1a). Therefore, the committee members in the generalized model are assigned to test samples in a dynamical manner.

It is reasonable to assume that an individual's signals may carry similar patterns to other persons in a specific context when they are performing a similar task. For a particular person for whom we want to build an engagement assessment model, if we can find similar patterns in the validation data (with labels) and select the classifiers that perform the best on the validation data, we may obtain good results by using these selected classifiers. This technique is different from the similarity-based individualization technique, which is based on a static selection procedure. Comparing to the similarity-based approach, we utilized a dynamic ensemble selection framework to dynamically individualize a generalized model to adapt to an individual's characteristics. The generalized model is built using available training datasets, which consists of a set of neural network classifiers as committee machine members. In dynamic ensemble selection, each classifier's accuracy is estimated in the local feature space surrounding the data similar to an unknown test sample from the individual. The first few top classifiers are then selected to classify the test sample by majority voting (Giacinto, 2001). The rational of this design is based on the assumption that if the test sample is similar to its local validation samples in the feature space, we may achieve a good assessment result for the test data point by utilizing those classifiers which perform well on the adjacent validation data points.

Dynamic ensemble selection tries to select the best set of classifiers in the committee that can optimally classify a given pattern. We can either select the best single classifier or a set of classifiers (ensemble) based on a validation data set to classify a given test sample. However, since data samples from a different feature space are, in general, associated with different classification difficulties, it is reasonable to expect better results from the selected ensemble rather than the best single classifier.

Figure 3 shows details in the feature space. For the given test sample represented by the red cross, a set of nearest neighbors (blue circles) are found in the validation dataset. Note that there are several different decision boundaries (blue curves) formed by the available trained classifiers. Those blue circles are then used to evaluate all classifiers and a set of best classifiers will then be selected to classify the test sample.

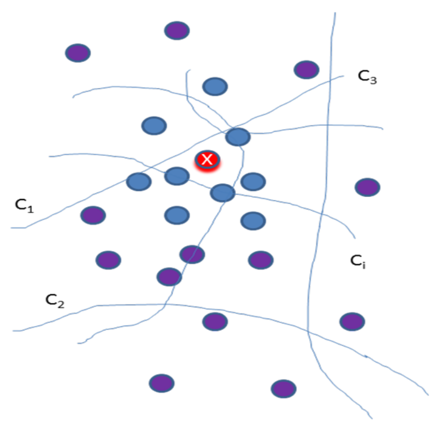

Figure 2. A detailed illustration for ensemble selection.

3 ENGAGEMENT ASSESSMENT EXPERIMENT DESIGN

In order to study engagement, we conducted experiments in a fully equipped Boeing 737 simulator involving commercial pilots (Ellis, 2009). The functionality of the simulator can be described as a fully functional flight deck with full glass cockpit displays, five outside visual projectors, functioning mode control panel with autopilot and autothrottle, and standard Boeing 737 controls. Several subjects participated in the pilot engagement study. Pilots had varying levels of experience with different types of aircraft. The experiment involved a flight from Seattle Tacoma International Airport to Chicago O'Hare International Airport. The details of the flight have been extracted from an actual American Airlines flight which took place on May 10th, 2010 (Zhang, et al., 2011). Details can be retrieved from on flightaware.com.

The experiment procedure has been previously described in our paper (Zhang et al., 2011). In order to study engagement, all pilots were scheduled to arrive at 5:30pm and were asked to avoid drinking caffeinated beverages such as coffee on the day of the experiment. An orientation video was shown to the subjects before the simulated flight. The video contained a description of the experiment as well as a Control & Display Unit (CDU) programming training section. The video included a description of the sensors and video recording devices used, as well as the responsibilities that the pilots would have during the experiment. The details shared with the subjects did not include information on the probes used to measure engagement levels so that pilots would not anticipate these probes throughout the experiment. During the flight simulation, one of the staff controlled the simulation computer to play pre-recorded audio files mimicking ATC transmissions. An experimenter was in charge of tagging the data to make sure that proper labels were added to the data sheets to identify the phases of the experiment as well as the times when the pilot responded to ATC. At the end of the experiment, the subjects filled out a subjective survey (such as NASA TLX) to assess their workload, fatigue and situational awareness during different phases of flight (Zhang, 2011).

In addition to the subjective surveys, objective data was collected from flight technical data (altitude, speed, *etc*.); three types of psycho-physiological sensors, including eye tracking cameras, an electroencephalogram (EEG) net, and electrocardiogram (ECG); and performance data such as response time to ATC calls or pump failure.

4 EXPERIMENTAL RESULTS

The developed techniques were evaluated with the experimental data collected through a Boeing 737 flight simulator. In this study, two engagement states and their time durations were first identified by evaluating videos of subjects. A pilot's state during takeoff or while handling a pump failure was considered as 'engaged' The pilot's state during level flight without any manipulation or if napping was recognized as 'disengaged'. Calculated features can then be labeled with these states by aligning with the identified time information.

We focused on the EEG sensor signals in this research. The EEG data was first preprocessed, including removal of environmental and DC artifacts, removal of EEG datasets with unreasonable measurements based on standard deviation (such as 0 indicating no signal collected), selection of EEG channels of interest, identification of spikes/excursions/amplifier saturation, removal of eye blink/body movement induced artifacts, and calculation of Power Spectrum Density (PSD) (Li et al, 2011). Based on existing research studies on EEG (Berka et al. 2007; Trejo et al., 2007) we selected a subset of EEG signals, including Fz-Oz, Cz-Oz, C3-C4, F3-Cz, F3-C4, Fz-C3, P7-Oz and P8-Oz. In this study, EEG absolute PSD variables for each 1-s epoch were computed and for each bipolar pair, the power spectrum within each band was summed up as a feature. All the features were analyzed and selected based on the PLOFS feature selection algorithm (Li et al., 2006) before being fed into the committee machine.

4.1 Similarity-based Model Individualization Evaluation

To study the effectiveness of the similarity-based model individualization technique, we first investigated the effectiveness of each similarity measure: t-test, entropy, Bhattacharyya distance, ROC and Wilcoxon test using the function provided in MATLAB 7.11 (the MathWorks Inc., Natick, MA, 2010). The hypothesis is that, if we can identify a similar subject in the training data pool, we can use his/her data to build a model and achieve a good performance when applying to the test subject.

We have evaluated the similarity–based individualization technique on 8 subjects. In Table 1, each row shows the testing performances for a subject using a model trained by the dataset of a subject indicated in the column. For example, the number shown in red represents the testing accuracy on subject 2 using the model trained by data from subject 3. This table will give us an overall cross subject performances.

Table 1. Performance (accuracy) summary for cross subject testing

Models from \ Subject	1	2	3	4	5	6	7	8
1	100	97.93	75.15	82.48	97.88	67.17	49.31	71.77
2	92.52	99.8	89.35	73.69	98	44.53	50.2	72.9
3	61.84	90.9	99.7	72.37	95.24	38.26	27.9	54.86
4	90.65	76.15	74.85	99.09	90.92	61.65	24.71	67.61
5	93.61	99.42	88.76	82.14	100	54.63	48.78	76.11
6	45.79	5	22.78	54.99	21.8	100	67.73	46.9
7	48.75	79.49	26.04	23.38	46.5	42.94	100	66.01
8	85.2	96.77	62.72	74.53	95.5	72.68	85.94	99.24

For each subject to be evaluated, we ranked all available subjects in the pool based on the cross subject performance and all of the five similarity measures. Among the top 2 ranked subjects by the cross subject performance, if they are the same as those ranked by a similarity measure, the similarity measure receives 2 point credits. If there is only one in common, the similarity receives 1 point. Also, we investigated the similarity measures by just using the features from a single class, i.e., engaged or disengaged. As an example, the following table shows the points that all the similarity measures have achieved under disengaged. A larger number means a better similarity measure. Similar observations can be achieved for the engaged state.

Table 2 Comparison of similarity measures

Subject	2	4	5	6	18	20	21	22	Sum
t-test	1	1	0	0	0	1	0	0	3
Entropy	1	1	0	0	0	0	0	0	2
Bhattacharyya Distance	1	0	0	0	0	0	0	1	2
ROC	1	1	0	0	0	1	0	0	3
Wilcoxon	1	1	1	1	1	1	0	1	7

The results show that the Wilcoxon test is the best similarity metric to measure the subject similarity. We further use the data trained from the most similar subject and evaluate the engagement assessment performance and the results are shown below.

Table 3 Comparison of performance: Generalized model vs. similarity-based model individualization

Subject	1	2	3	4	5	6	7	8	Average
Generalized	93.98	99.28	74.17	83.8	98.24	63.2	43.75	80.98	79.67
Similarity-based	90.65	97.93	88.76	82.48	97.88	61.65	49.31	71.77	80.05

The performance by the model individualization does not show an improvement on the performance comparing to the performance achieved using a generalized model (trained using data from other subjects). Further investigation is needed with a large data corpus.

4.2 Dynamic Ensemble Selection for Model Individualization

We used eight subjects' data collected from the Boeing 737 flight simulator and evaluated the dynamic ensemble selection technique for model individualization. There are seven scenarios, as described below, to test the engagement assessment performances.

- Scenario 1: we utilized the first 20% of data from one subject for training (26 committee members in total) and the remaining 80% data for testing the same subject, thus obtaining the individual model performance.

- Scenario 2: Generalized model performance. For each subject, we trained 5 committee members/models using data from each of other subjects. Since there are 8 subjects in total, 5*7=35 committee members were trained for the testing subject. This will give us the baseline generalized model performance.
- Scenario 3: The same as Scenario 2 except that the dynamic ensemble selection technique was applied using data from other subjects for validation.
- Scenario 4: The same as Scenario 3 except that the validation dataset for dynamic ensemble selection was from the testing subject (first 20%).
- Scenario 5: For each testing subject, there were committee members trained using the testing subject's first 20% of the data and another 5*7 models trained using data from other subjects.
- Scenario 6: The same as Scenario 5 except that we used the dynamic ensemble selection technique and combined the data from other subjects and the first 20% of data from the testing subject as the validation dataset.
- Scenario 7: The same as Scenario 6 except that the validation data only contained the first 20% of dataset from the testing subject.

The 7 scenarios are summarized in Table 4and the performance of engagement assessment for all the subjects are shown in Table 5.

Table 4 Experiment setup

Scenario	No. of Models	Ensemble Method	Validation dataset
1	26	Majority vote	N/A
2	5*7	Majority vote	N/A
3	5*7	Dynamic	From all other subjects
4	5*7	Dynamic	Top 20% of data from the testing subject
5	26 + 5*7	Majority vote	N/A
6	26 + 5*7	Dynamic	From all other subjects
7	26 + 5*7	Dynamic	Top 20% of data from the testing subject

Table 5 Experiment results

Scenario / Subject	1 (Individual Model)	2 (Generalized Model)	3 (Dynamic Ensemble Selection)	4 (Dynamic Ensemble Selection)	5 (Individual Model)	6 (Dynamic Ensemble Selection)	7 (Dynamic Ensemble Selection)
2	88.74	93.98	89.51	96.31	94.56	91.26	92.82
4	96.12	99.28	96.55	97.99	97.55	96.26	96.98
5	86.35	74.17	79.7	85.24	85.98	81.55	90.04
6	90.51	83.8	79.27	86.24	93.21	85.37	91.29
18	92.84	98.24	94.82	97.36	97.47	91.85	96.04
20	99.53	63.2	75.82	82.48	94.86	81.54	99.65
21	94.41	43.75	61.04	84.44	87.9	74.87	94.41
22	88.09	80.98	79.48	80.66	88.33	84.32	87.38
Average	92.07	79.67	82.02	88.84	92.48	85.87	93.57

We found that 20% of the data from each subject is sufficient to train a reasonably good model for engagement assessment (Scenarios 1 & 5), which performs better than the generalized model (Scenario 2). The dynamic ensemble selection strategy boosts the performance of the model even without data from the individual (Scenario 3). If we use data from a test subject to dynamically select committee members, the performance can be further improved (Scenarios 4, 6 and 7).

5 DISCUSSIONS

In this research, we have explored two different model individualization techniques for engagement assessment in aviation environments. Future tasks include enhancing the model with additional sensory information (flight technical data and ECG, for example), further improving the model individualization technique and continuing to further verify and validate the real-time assessment technique with additional participants' data.

ACKNOWLEDGMENTS

This project was funded by the NASA (Contract No: NNX10CB27C). We thank Dr. Alan T. Pope, Mr. Chad L. Stephens, and Dr. Kara Latorella for their comments and suggestions as we performed this research.

REFERENCES

C. Berka, et al., "EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks," Aviation, Space and Environmental Medicine, Vol. 78, No. 5, Section II, May 2007

W. Boucsein, R. W. Backs, "Engineering psychophysiology as a discipline: historical and theoretical aspects," in Boucsein, W., Backs, R.W. (Eds.), Engineering Psychophysiology: Issues and Applications. Erlbaum, London.

Breiman, L., Bagging Predictors, Machine Learning, 24(2), pp.123-140, 1996

K. Ellis, "Eye Tracking Metrics for Workload Estimation in Flight Deck Operations, Masters Thesis, University of Iowa, 2009.

Engel, B.T., Stimulus-response and individual-response specificity. Archives of General Psychiatry, 1960. 2: p. 305-313.

Foerster, F., Psychophysiological response specificity: A replication of a 12 month period. Biological Psychology, 1985. 21: p. 169-182.

Giacinto G, Roli F. (2001): Dynamic Classifier Selection Based on Multiple Classifier Behaviour. Pattern Recognition 34, 179-181

Robert Hockey, "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," NATO ASI SERIES Vol 355

F. Li et al., "Engagement Assessment Using EEG signals," MODSIM World Conference, 2011, Virginia Beach, VA

J. Li, J. Yao, N. Petrick, R. M. Summers, and A. K. Hara, "Hybrid Committee Classifiers for a Computerized Colonic Polyp Detection System," Medical Imaging 2006: Image Processing, vol. 6144, pp. 1701-1709, 2006.

Marwitz, M. and G. Stemmler, On the status of individual response specificity. Psychophysiology, 1998. 35(1-15).

Erik Olofsen, Hans P A Van Dongen, Christopher G Mott, Thomas J Balkin, David Terman, "Current Approaches and Challenges to Development of an Individualized Sleep and Performance Prediction Model" The Open Sleep Journal (2010), Volume: 3, Pages: 24-43

A. Pope, E. H. Bogart, D. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," Biological Psychology; 40(1): 187-195. (1995)

Rajaraman, S., A. V. Gribok, N. J. Wesensten, T. J. Balkin, and J. Reifman. An Improved Methodology for Individualized Performance Prediction of Sleep-deprived Individuals with the Two-process Model. Sleep. 2009 October 1; 32(10):1377-1392

D. Schmorrow & A. Kruse, "Improving human performance through advanced cognitive system technology. Interservice/ Industry Training," Simulation and Education Conference, Orlando, FL; 2002.

D. Schmorrow, K.M. Stanney, G. Wilson, and P. Young, "Augmented cognition in human-system interaction," In: Salvendy G, ed. Handbook of human factors and ergonomics, 3rd ed. New York: John Wiley; 2005.

M.E. Smith, L. McEvoy, and A. Gevins, "Neurophysiological indices of strategy development and skill acquisition," Cognitive Brain Research, Vol. 7, issue 3, Jan. 1999, pp. 389– 404

R. H. Stevens , T. Galloway, and C. Berka, "EEG-Related Changes in Cognitive Workload, Engagement and Distraction as Students Acquire Problem Solving Skills," Proceedings of the 11th international conference on User Modeling. 2007

Leonard J. Trejo et al., "EEG-based Estimation of Cognitive Fatigue," Proceedings of SPIE: Bio-monitoring for physiological and cognitive performance during military operations. v5797. 105-115.

G. Zhang et al., "Individualized Cognitive Modeling for Closed-Loop Task Mitigation, ModSim World Conference, 2009, Virginia Beach, VA

G. Zhang et al., "A Systematic Approach for Real-Time Operator Functional State Assessment," MODSIM World Conference, 2011, Virginia Beach, VA