31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Recombination of Common Sensory-Motor Impairment Evaluation Techniques using a Committee of Classifiers

Nathan Chalmers, Geoffrey Seaborn, Jae-Yoon Jung, Janice I. Glasgow, and Stephen H. Scott

Abstract— **Conventional methods for assessing levels of sensory-motor impairment in stroke patients are inherently subjective and dependent upon the clinician's own observations and opinions. In this study, 93 control and 63 stroke subjects underwent robotic assessment to gauge sensory-motor impairment. Multiple statistical data normalization and dimensionality-reduction measures were evaluated, using four different classifier types, in order to derive an optimal feature vector for the purpose of distinguishing stroke from control subjects. The optimal feature vector was then utilized to train a committee of classifiers for the purpose of recombining data from several traditional sensory-motor assessment scores into a KINARM specific assessment metric. We were able to create a training vector capable of distinguishing between stroke and control subjects with high accuracy, and demonstrated that the committee of classifiers assigned consistent scores to patients of similar levels of impairment.**

I. INTRODUCTION

Stroke is a loss of function caused by an interruption in blood flow to the brain. This shortage can result in paresis, loss of sensation, loss of vision, loss of memory and the inability to speak or understand speech [2]. In order to achieve a successful recovery after stroke, it is necessary to understand the motor impairments underlying the patient's condition [10]. Although the effects of a stroke are both severe and sudden, most stroke survivors recover to some degree [1]. When patients are carefully assessed, and an optimal rehabilitation strategy is selected, outcome can be optimized for recovery speed and physical independence.

Methods used for assessing and determining an appropriate type of therapy for each stroke patient are inherently subjective as they depend largely on the clinician's own observations, evaluations, and opinions [4]. Rating systems are mostly based on coarse numerical grading, and do not provide therapists or physicians with information about the detailed kinematics underlying the motor deficits [1].

The KINARM (Kinensiological Instrument for Normal and Altered Reaching Movements, BKIN Technologies, Kingston, ON) [12] is a robotic exoskeleton developed for the purpose of sensing and perturbing limb motion. Patients attached to the device are capable of flexing and extending both shoulder and elbow joints allowing them to move their hand within the horizontal plane. Hand speed and position within the plane are monitored using a series of torque motors and accelerometers [11]. Through continuous monitoring of a patient during a variety of reaching tasks, the device is capable of quantifying even subtle degradations in sensory and motor faculties at a level of precision not perceivable to human eye [1]. The development of computational tools derived from this precise motor-sensory data has the potential to improve upon patient quality of life by increasing assessment accuracy and thereby aiding in the planning of rehabilitation therapies.

In our study, we used statistical approaches and information-theoretic measures to isolate promising features from KINARM experimental data for the purpose of classification. We then used the output from a committee of classifiers to create a KINARM-specific stroke impairment assessment score. Our study was comprised of two parts:

The first part involved isolating descriptive features from experimental session data, for the purpose of distinguishing stroke and control subjects, using various dimensionality-reduction and data-normalization techniques. A performance comparison between K nearest neighbor (KNN) [3], support vector machine (SVM) [9], artificial neural network (ANN) [8], and linear discriminate analysis (LDA) classifiers [13] was conducted to determine both the optimal training vector and classifier modality.

The second part of our study involved the creation of a committee of classifiers for the purpose of recombining information gained from traditional assessment scores into a unique sensory-motor impairment assessment metric.

It is our belief that through the utilization of several classification and feature extraction approaches it is possible produce an accurate, consistent, and reliable metric for evaluating sensory-motor impairment in stroke patients.

II. METHODS

A. Subjects and Task

55 ischemic stroke and 8 hemorrhagic stroke patients with a median age of 66 years and 93 control patients with no known prior history of sensory-motor impairment and a median age of 61 years were selected for robotic assessment using the KINARM. One of eight virtual targets were projected onto a two-dimensional virtual reality display. Upon signaling from an indicator light, subjects were instructed to move their hand from the center point to the projected peripheral target 10 cm away. The reaching task was defined as being complete when the subjects hand entered and remained within the bounded target area. A total of 8 attempts were made for each of the 8 peripheral targets

Nathan Chalmers, Geoffrey Seaborn, Jae-Yoon Jung and Janice I. Glasgow are with the School of Computing, Queen's University, Kingston, Ontario, Canada (email: {nathan, geoff, jung, janice}@cs.queensu.ca).

Stephen H. Scott is with the Department of Anatomy and Cell Biology, Canadian Institute of Health Research Group in Sensory-Motor Systems, Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada (email: steve@biomed.queensu.ca)

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

yielding 64 potential trial sets in the event the subject was able to complete all trials.

Fourteen metrics describing path length and hand speed during various portions of the reaching task were calculated based on this experimental data. FMTDist, FMTmaxSP, FMTDirErr, and FMTDistErr refer to the distance traveled, maximum speed, the angle between the optimal trajectory and the patient's trajectory, and the distance of the hand from the target during the initial portion of the reaching task. TotalMT, PathLen, MTMaxSP, and NumMTmaxSP refer to the total movement time, path length, maximum speed, and the number of speed maxima throughout the course of the entire reaching task. PathLenRatio and FMTMaxSPRatio refer to the ratio between subject path length and optimal path length, and the ratio between the maximum speed achieved during the initial hand movement and total hand movement.

A unique session vector was created for each subject by calculating the median of each metric across all trials. Session vectors were only calculated from trials involving the affected hand in stroke subjects and the dominant hand in control subjects.

Within a week of robotic assessment, clinicians assessed sensory-motor impairment of each subject, and the measures include the Purdue Pegboard score (Purdue), dynamometer score for hand squeeze (DynaHand), dynamometer score for hand pinch (DynaPinch), Functional Independence Measure (FIM), Chedoke-McMaster score of the arm (CMArm), Chedoke-McMaster score of the hand (CMHand), and thumb proprioception score (Thumb).

B. Data Normalization and Dimensionality-Reduction

Four stroke and control training sets were created from the combination of two distinct normalization and dimensionality-reduction methods. Normalization method 1, herein referred to as $N1$, consisted of calculating the base 2 log of the data set, followed by centering and scaling each KINARM metric to produce a mean of 0 and a standard deviation of 1:

$$z_{N1} = \frac{log_2(x) - \mu_{metric}}{\sigma_{metric}} \tag{1}$$

The second normalization method, herein referred to as $N2$, utilized the maximum and minimum value of each KINARM metric to scale the data between 0 and 1.

$$z_{N2} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Dimensionality-reduction and feature selection methods pose several practical advantages through alleviating the curse of dimensionality, streamlining the learning process, and by improving generalization through the selection of only relevant features [5]. It is in this regard that a statistical filter to remove consistent variables between control and stroke subjects, and two distinct dimensionality-reduction techniques, were employed to reduce the fourteen KINARM metrics to eight.

TABLE I
KOLMOGOROV-SMIRNOV TEST P-VALUES

Metric	P-Value	Metric	P-Value
FMTDirErr	4.778e-36	FMTDist	3.435e-36
FMTDistErr	3.161e-35	FMTDistRatio	7.034e-58
FMTmaxSP	1.122e-40	FMTMaxSPRatio	4.275e-95
MinMaxSPDiff_mean	6.584e-35	MTMaxSP	7.567e-41
NumMTmaxSP	1.1297e-118	postureSPp50	5.802e-35
PathLen	1.637e-40	PathLenRatio	1.429e-99
RT	1.747e-47	TotalMT	4.761e-79

TABLE II
SPEARMAN RHO CORRELATION COEFFICIENTS

Metric	Correlated Metric	Corr. Coeff.
PathLen	**PathLenRatio**	0.8916
FMTDist	**FMTmaxSP**	0.7751
FMTDist	**FMTDistRatio**	0.749
FMTmaxSP	**MinMaxSPDiff**	0.6769
MinMaxSPDiff	PathLenRatio	0.6653
FMTDistErr	**NumMTmaxSP**	0.6061

A Kolmogorov-Smirnov test was performed at the 5% level of significance to judge metric conformance to the normal distribution. No KINARM metrics were found to be normally distributed as shown in Table I and non-parametric tests were used for all statistical analysis as a result.

A non-parametric Man-Whitney U-Test [6] was performed at the 5% level of significance to isolate KINARM metrics that did not significantly differ between the control and stroke subject groups. FMTMaxSPRatio ($p = 0.7953$) and MTMaxSP ($p = 0.0952$) were found not to significantly differ between control and stroke subject groups, and were subsequently removed from the training set.

Dimensionality-reduction method 1, herein referred to as $D1$, involved the calculation of Spearman's rank-correlation coefficient for all KINARM metrics in a pair-wise manner on a combined stroke and control subject data set [15]. In Table II the six pairs of KINARM metrics exhibiting the highest Spearman Rho Correlation Coefficient are shown. The first eight distinct KINARM metrics associated with highest correlation coefficients, highlighted in bold, were chosen to form the training set.

Dimensionality-reduction method 2, herein referred to as $D2$, consisted of performing principle component analysis on the normalized data set, followed by selecting the first eight principle components of both the stroke and control subject groups [7].

C. Classification Task

ANN, SVM, KNN, and LDA classifiers were trained on the four normalized and reduced data sets $(D1, N1)$, $(D1, N2)$, $(D2, N1)$, and $(D2, N2)$. ANN structure consisted of a three layer feed forward network composed of 8 input nodes, a varying number of h hidden nodes, and two output nodes representing class control and stroke respectively. ANNs were trained using a standard back-propagation training algorithm utilizing gradient-descent with a mean squared error cost function [3]. The KNN classifiers were built and trained utilizing a cosine coefficient similarity metric. The optimal number of nodes h in the hidden layer and number K for KNN classifiers was determined by empirical testing between the values 1 and 25. The values which yielded the highest mean-percent classification accuracy were selected. This was determined to be h = 14 for the ANN and K = 9 for the KNN classifier. SVMs were trained using a linear kernel and 1-norm soft margin optimization algorithm. LDA classifiers were trained through fitting a multivariate normal distribution to the data with a pooled estimate of covariance.

D. Committee of Classifiers

A committee of classifiers was constructed to combine knowledge regarding clinical sensory-motor impairment assessment scores into a KINARM-specific sensory-motor impairment score. For each of the seven clinical assessment scores, a committee member was trained to distinguish between moderately-impaired and severely-impaired stroke subjects. The median value of each clinical assessment score amongst its members was determined and used as a partition value, as shown in Table III. Stroke subjects with assessment values less than the threshold value were designated severely-impaired, those above the threshold value were designated moderately-impaired.

Support vector machines were chosen to form the individual committee members as a result of the ANN's tendency to produce binary results. The (D1, N2) normalization reduction method was found to yield the greatest mean-percent classification accuracy during the classification task, and was subsequently chosen to reduce and normalize committee member training sets. Committee member training data was limited to stroke subjects for which the committee member's clinical assessment score was known. In a manner identical to the classification task, each SVM was evaluated using 10-fold stratified cross-validation in order to gauge effectiveness. Mean-percent classification accuracy and the Shannon entropy, calculated using maximum likelihood estimates derived from a confusion matrix, were recorded and shown in Table III.

A linear combination of committee member outputs was used to determine the KINARM-specific assessment score. For each sample vector s, and committee member SVM_i, with normal vector w_i and bias γ_i, (3) was calculated to produce x_i. The results of (3) for each committee member were summed, as illustrated in (4), to produce a KINARM assessment score r. Both the mean-percent accuracy and the Shannon entropy of each committee member were used as c_i in separate trials.

$$x_i = s \cdot w_i - \gamma_i \tag{3}$$

$$r = \sum_{i=1}^{n} c_i x_i \tag{4}$$

TABLE III
COMMITTEE MEMBER THRESHOLD VALUES AND COMBINATION COEFFICIENTS

Assessment Score	Threshold Value	Accuracy	Entropy
Purdue Pegboard	4.816	50.00	1.288
DynaHand	15.89	52.50	1.254
DynaPinch	5.631	53.34	1.366
FIM Score	80.84	50.91	1.378
CMArm	5.483	53.33	1.352
CMHand	5.29	63.33	1.156
Thumb	0.639	60.00	1.119

TABLE IV
CLASSIFICATION ACCURACY FEED FORWARD ANN WITH h = 14

Data Set	Mean	Maximum	Minimum
$(D2, N1)$	64.77	77.42	54.84
$(D2, N2)$	68.25	87.10	48.38
$(D1, N1)$	82.32	100.0	58.06
$(D1, N2)$	94.45	100.0	87.10

III. EXPERIMENTAL RESULTS

A. Classification Results

10-fold stratified cross-validation was employed to evaluate each of the classifiers on the four normalized and reduced data sets [14]. Elements within the data sets were randomly-ordered, and the cross-validation process was repeated five times. Mean-percent classification accuracy, maximum classification accuracy, and minimum classification accuracy were recorded. Dimensionality-reduction, normalization strategy $(D1, N2)$ was found to achieve the greatest mean-percent classification accuracy amongst all four classifiers. When trained on data set $(D1, N2)$ the ANN was found to achieve the greatest mean-percent classification accuracy of the four classifiers.

B. Committee Results

Eight of the sixty three stroke subjects had multiple pieces of KINARM session data recorded within a twenty four hour period. For each stroke subject with multiple pieces of KINARM data, a committee of classifiers was trained utilizing the remaining stroke subjects. As summarized in Table VI, committee results were calculated for each piece of multiple session data using both the mean-percent classification accuracy and Shannon entropy as a combination coefficient. A Mann-Whitney U-Test was performed at the 5% level of significance to determine whether committee scores significantly differed between multiple pieces of session data

TABLE V
CLASSIFICATION ACCURACY OF A KNN WITH k = 9, SVM, AND LDA CLASSIFIERS

Data Set	KNN	SVM	LDA
$(D2, N1)$	65.80	57.42	34.83
$(D2, N2)$	73.55	56.77	37.41
$(D1, N1)$	81.29	54.84	39.35
$(D1, N2)$	92.25	92.90	91.90

TABLE VI

COMMITTEE SCORES FOR PATIENTS WITH MULTIPLE PIECES OF SESSION DATA

Subject	$c_i = PercentAccuracy$		$c_i = ShannonEntropy$	
	Session 1	Session 2	Session 1	Session 2
1	1.8600	1.8375	4.4787	4.4718
2	-2.6265	-3.8458	-6.5578	-9.3358
3	-2.5716	-1.71522	-6.0305	4.0081
4	-0.1955	-1.9324	-0.5495	-4.5282
5	0.0539	-3.2723	0.2455	-7.8502
6	0.2866	1.7006	0.5648	3.7574
7	1.4992	0.5179	3.914	1.4178
8	-0.4630	0.5073	-0.9632	1.13196

for each stroke subject. Scores produced by the committee using mean-percent classification accuracy $(p = 0.8785)$ and Shannon entropy $(p = 0.7984)$ combination coefficients were found not to significantly differ between multiple pieces of session data. Scores produced by the committee using mean percent classification accuracy and Shannon entropy combination coefficients were both found to have a Spearman rho correlation coefficient of 0.7619 $(p = 0.0280)$.

IV. DISCUSSION

As illustrated in Table III and Table IV, the data set that produced the highest mean-percent classification accuracy amongst all four classifiers was that in which the dimensionality was reduced using Spearman correlation and normalized by scaling the KINARM metrics between 0 and 1. Principle component analysis, a naive form of feature extraction, failed to produce results as accurate as those obtained from the Spearman correlation reduced set when implemented with the same normalization technique.

Though this does not demonstrate that the Spearman correlation dimensionality-reduction method is capable of producing a globally-optimal training vector, it does demonstrate that it is optimal in the given context of differentiating stroke and control subjects. The assumption is then made, that given local optimality in the control-versus-stroke classification task, the dimensionality-reduction method may serve as a good candidate for the creation of training vectors for similar problems. It was from this reasoning that the Spearman correlation dimensionality-reduction and scaling normalization methods were chosen to process raw KINARM data for the committee of classifiers.

As previously stated, the committee results have been demonstrated to not differ significantly $(p = 0.8785, p = 0.7619)$ and have strong correlation $(\rho = 0.7619, p = 0.0280)$ between multiple pieces of session data for both mean-percent-classification accuracy and Shannon entropy classification coefficients. Since statistical consistency and strong correlation have been shown within individual subjects, where it is assumed that the level of sensory-motor deficit does not improve between sessions, it may be possible to infer consistency and strong correlation between different patients with similar levels of impairment.

V. CONCLUSION

The classifiers trained to differentiate between control and stroke subjects were found to have a high level of accuracy when trained using $(D1, N2)$ dimensionally-reduced and normalized training sets. The committee of classifiers constructed using the $(D1, N2)$ dimensionality-reduction and normalization method was found to be capable of generating consistent results amongst subjects with multiple pieces of KINARM session data.

With the successful derivation of a consistent KINARM-unique assessment score, we feel we have created a sensory-motor assessment system that could be useful for planning effective rehabilitation therapy for stroke victims. Future work will concentrate on the proper scaling and formalization of the KINARM-unique assessment score for the chance that it, like the KINARM itself, may be suitable for wide-scale clinical use.

VI. ACKNOWLEDGMENTS

We thank K. Moore and H. Bretzke for technical help. This work was supported by NSERC and Canadian Institutes of Health Research (CHIR) MOP 81366.

REFERENCES

[1] A. Abu-Zeid, *Parameter analysis for robotic assessment of impairments in reaching due to stroke*, M.Sc. Thesis Queen's University 2007.

[2] H. P. Adams, V. Hachinski, and J.W. Norris, *Ischemic Cerebrovascular disease*, Oxford University Press 2004.

[3] N. Bryant, G. Hepner, T. Hepner, and N. Ritter, "Artificial neural network classification using a minimal training set - Comparison to conventional supervised classification", *Photogrammetric Engineering and Remote Sensing*, 56:469-473, 1990.

[4] D. J. Gladstone, C. J. Danells and S. E. Black, "The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties", *Neurorehabilitation and Neural Repair*, 16 (3):232-240, 2002.

[5] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", *The Journal of Machine Learning Research*, 3:1157-1182, 2003.

[6] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, Wiley-Interscience 1999.

[7] I. Jolliffe, *Principal Component Analysis*, Springer 2005.

[8] J. Laaksonen, "Classification with learning k-nearest neighbors", *IEEE International Conference on Neural Networks*, 3:3-6, 1996.

[9] J. Li, R. Janardan and Q. Li, "Two-Dimensional Linear Discriminant Analysis", *Advances in Neural Information Processing Systems*, 17:1569–1576, 2004.

[10] S. Michaelsen, "Short-term effects of practice with trunk restraint on reaching movements in patients with chronic stroke: a controlled trial", *Stroke*, 35:1914-1919, 2004.

[11] S. H. Scott, "Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching", *Journal of Neuroscience Methods*, 89:119-127 1999.

[12] K. Singh and S. H. Scott, "A motor learning strategy reflects neural circuitry for limb control", *Nature Neuroscience*, 6 (4):399-403, 2003.

[13] J.A.K. Suykens and J Vandewalle, "Least Squares Support Vector Machine Classifiers", *Neural Processing Letters*, 9 (3):1370-4621, 2004.

[14] X. Zeng, "Distribution-balanced stratified cross-validation for accuracy estimation", *Journal of Experimental and Theoretical Artificial Intelligence*, 12 (1):1-12, 2000.

[15] J. H. Zhar, "Significance Testing of the Spearman Rank Correlation Coefficient", *Journal of the American Statistical Association*, 67 (339): 578-580, 1972.