DETECTING MILD COGNITIVE LOSS WITH CONTINUOUS MONITORING OF MEDICATION ADHERENCE

Yonghong Huang, Deniz Erdogmus, Zhengdong Lu, Todd K. Leen

Department of Computer Science & Electrical Engineering
Oregon Health & Science University, Portland, OR, USA

ABSTRACT

This paper describes an approach for detecting early cognitive loss using medication adherence behavior. We investigate the discriminative power of a comprehensive set of recurrent medication timing features extracted from time-of-day and inter-dose timing statistics. We adopt information theoretic measures for feature ranking for initial dimensionality reduction and conduct exhaustive leave-one-out cross validation for final feature selection and regularization. The selected feature set is subjected to a support vector machine for classification. The results demonstrate that patterns of adherence based on the data from relatively unobtrusive behavior monitoring can make reliable inference for mild cognitive loss individuals.

Index Terms— cognitive loss detection, medication adherence, continuous monitoring, pattern recognition

1. INTRODUCTION

Medication adherence can be a complex cognitive behavior for elderly due to age-related declines in cognitive functioning [1]. Recently Hayes and her colleagues investigated a group of independently-living elderly and demonstrated that early cognitive changes contributed to lapse of medication adherence [2]. The goal of our research is to develop algorithms for detecting changes of cognitive decline for aging people based on their capacity to comply with medication regimens. Our initial results showed that medication adherence behavior could be a cue to detect mild cognitive loss individuals from cognitive health individuals [5]. In this study, we explore a complete set of features and apply an alternative classifier - support vector machine (SVM) to the task of discriminating between healthy individuals and those with early cognitive loss, on the basis of medication adherence behavior. The results demonstrate that patterns of adherence based on the data from relatively unobtrusive behavior monitoring can be a reliable detector for mild cognitive loss individuals.

2. COHORT AND DATA

Forty independently-living elder subjects were recruited for the study [2]. All had baseline Mini-Mental State Examination (MMSE) scores greater than 24, and Clinical Dementia Rating (CDR) of 0 or 0.5. Subjects were divided into two groups based on their memory function as assessed by Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) scores. Graham's normative ADAS-Cog data [4] were used to generate an age-adjusted 95% confidence interval for cognitively healthy individuals.

We define the *High Cognitive Performer* (HCP) group as those subjects whose age-adjusted ADAS-Cog scores fell within this interval (N=19), and the *Low Cognitive Performer* (LCP) group as those subjects whose scores fell outside this interval (N=21). After the data collection, results from two individuals were removed for this study: data from one HCP subject was corrupted by equipment failure, and one LCP subject met the clinical standard for mild cognitive impairment (MCI), greater cognitive loss than the intended target population. This left 38 subjects, N=18 in the HCP group and N=20 in the LCP group.

Subjects were instructed to take vitamin C tablets twice daily at agreed-upon times, one in the morning, and one in the evening. Their behaviors were monitored by an in-home device, called MedTracker, developed by Hayes et al [3]. The device consists of a seven-day pillbox instrumented to record the time of opening of the compartments. Event times are stored in an on-board buffer and transferred from the device by Bluetooth wireless every two hours.

Subjects were monitored for approximately five weeks. A sample time series of events (from the HCP group) recorded by the MedTracker is shown in Figure 1. Dots mark the compartment openings, and crosses mark the missed doses. The two solid horizontal lines (at approximately 6:00am and 5:30pm) mark the *planned dose times* (explained below) for this subject. The dashed horizontal lines bound a window one hour before and two hours after the planned dose times.

3. CLASSIFIER CONSTRUCTION

The classification study is to assess the MedTracker as an indicator of cognitive health. The goal is to select the most discriminatory feature combinations of medication adherence and evaluate the efficacy of SVM classifier on the discrimination of the *LCP individuals* and the *HCP individuals*. We adopt the classification accuracy to estimate the quantitative efficacy.

3.1 Data Preparation

Three aspects of the subjects' dose-taking behavior created the need for some care in the data analysis. First, although instructed to take one dose in the morning and one in the afternoon, the noon hour was not a reliable boundary between events for each subject due to non-adherence in the data. To determine an appropriate boundary between the two sets of events for each subject, we clustered the subject's events into morning and evening groups using the Matlab function *clusterdata*. The algorithm returned a good partition except on two subjects with outliers. We manually corrected the clustering by excluding the outliers when doing the clustering and then restoring those outliers back to one or the other group based on the visual inspection of the data. Second, ideally two doses in the morning and in the afternoon were in the same day (within 24 hours), but three subjects had several afternoon events cross midnights. To include these events within 24 hours, we shifted all of the events two hours early for these subjects. Third, although each subject agreed on the planned AM and PM times for their doses, their actual median dose times deviated considerably from their plan. For analysis we *define* a surrogate *planned time* for the AM dose as the median time for all morning events in a subject (rounded to the nearest half-hour) due to non-adherence. Planned time for the PM dose is similarly defined. The median times are shown by solid horizontal lines in the example time series in Figure 1.

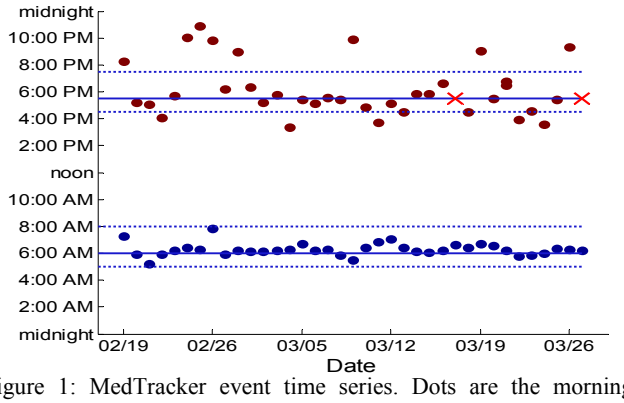

Figure 1: MedTracker event time series. Dots are the morning events and the afternoon events. Solid lines are the median of morning times and the median of evening times. Dash lines are the leeway of compliance of the set time (one hour before the set time and two hour after the set time). Crosses are the missing events.

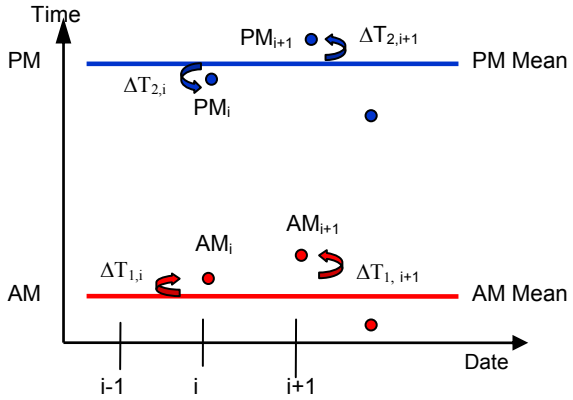

Figure2: An illustration of the inter-dose timing. The morning and evening event times for day *i* are denoted as AM_i and PM_i. The difference of the AM event of day *i* from the AM mean is denoted as $\Delta T_{1,i}$ and the difference of the PM event of day *i* from the PM mean is denoted as $\Delta T_{2,i}$.

3.2 Feature Extraction

The goal of the feature study is to evaluate the discriminate power of the features on the discrimination between the *LCP group* and the *HCP group*. We adopt the leave-one-out cross validation and information theoretic measures for feature selection. We apply the selected feature set to the SVM for classification.

3.2.1. Description of feature set

In the initial study [5], we summarized the time series by four features that describe the subject's dose behavior. We demonstrated that not all of the features are equally useful for discriminating between the two groups. In this study, we added features that captured the correlation statistics of inter-dose timing. Figure 2 shows the inter-dose timing. Feature1 is the percentage of the individual's events for which the dose is taken close to the planned time; no more than one hour before or two hours after the planned (that is, the median AM or PM) time; Feature2 is the percentage of days with no less than two compartment openings; Feature3 is the standard deviation of the time of the evening dose; and Feature4 is the standard deviation of the time of the morning dose. Features 5 and 6 are the standard deviation of the difference between corresponding two dose events. Features 7-10 are correlation features, which are the covariance measuring the nature of the association of two dose events. If they are statistically independent, the covariance is zero. Features 11-14 are correlation coefficients of the corresponding two events.

A complete list of the full feature set with the associated equations is shown in Table 1. Let N_{taken} denote the total number of the taken events as the planned time (defined in Section 2.3), N_{missed} denote the total number of the missed events and the taken events not as the planned time, $N_{prescribed}$ denote the total number of the events as the prescribed (at least two doses in a day), $N_{indates}$ denote the dates which are not out of town, *Std* denote the standard deviation and *E* denote the expectation operator. Below we will discuss how we choose the optimal subset of features to obtain good classification.

3.2.2 Reason for reduced cohort

The correlation features involved two-dose events, so we had to pre-process data to generate the correct features. First, we selected the last events when there were multi-

events happened in the morning or in the afternoon. We selected the dates when both events (morning event and afternoon event) happened in the same date and the date after and removed the dates with missing events. Second, we excluded three subjects containing less than three-day events from the *LCP* group because we needed at least three days with both events in the morning and in the afternoon to calculate correlation coefficients across days. Thus we had 35 subjects (20 *HCP* and 15 *LCP*).

3.2.3. Feature Selection

Due to the huge number of overall possible feature combination subsets, the exhaustive cross-validation evaluation strategy is not feasible. Consequently, we employ an information theoretic feature ranking and selection strategy [9] to identify a worthy subset of features and then conduct extensive leave-one-out cross-validation procedure [6,7] on the selected features. Specifically, we employ kernel density estimation and mutual information (KDE-MI) method to rank the features.

Feature set {1, 2, 3, 4, 10, 12, 13} were selected as the top feature subset for the exhaustive search for further reduction. To identify the subset of features with the most discriminative power, feature ranking is performed on the 14 features described in Table 1 using the mutual information based greedy search technique. Ranking with KDE-MI proposed features {1, 6, 2, 12, 8} as the top five. The validation results for individual feature showed that features {1, 2, 3, 4, 6, 10, 12, 13} are the top eight features. This information was combined to arrive at a final feature subset {1, 2, 3, 4, 10, 12, 13} for exhaustive cross-validation evaluation (We did not select features 6 & 8 because they contained similar information with features 10 & 12, based on the formulas in Table 1).

3.3 Classifier Description

Our goal in classification is to build a classifier that uses an appropriate combination of the features as input, and accurately assign individuals to the *LCP* or the *HCP* groups. We adopt SVM [7,8] as the classifier.

3.3.1 Support Vector Machine

A radial basis (Gaussian kernel) SVM is used in this study. The SVM is optimized to construct a maximum-margin separating hyperplane by mapping input vectors to a higher dimensional space. The separating hyperplane is the hyperplane that maximizes the distance between the two parallel hyperplanes on each side of the boundary touching the closest data (support vectors) from each class. The assumption is that the larger the margin between these parallel hyperplanes the less the generalisation error will be. A cost parameter C in the optimality criterion controls the number of support vectors and the trade-off between learning error (margin) and model complexity (the size of the slack variables). A larger C corresponds to assigning a higher penalty to errors (when the classes are not separable by a hyperplane in the feature space). To find the optimal hyperplane, the SVM is trained and optimized by solving a convex quadratic programming problem. After training, the optimal Lagrange multiplier for each sample and weights are obtained. The support vectors, which are the data points lying at the border of the margin have non-zero optimal solutions for their coefficients in the final discriminant, while others converge to zero weights, thus leading to a sparse nonparametric forward discriminant function.

The kernel size and the cost parameter C can be chosen by users. However, to avoid overfitting, we adopt a leave-one-out cross-validation to adjust these regularization parameters, the kernel size σ and the cost parameter C. The leave-one-out procedure is conducted on the training session to select the optimal parameters and the parameters are then applied to the *independent* test data to do the classification.

Table 1: A list of the complete feature set

Feature Number	Feature Name	Equation
1	As planned	$\frac{N_{taken}}{N_{taken}+N_{missed}}\times 100$
2	As prescribed	$\frac{N_{prescribed}}{N_{indates}}\times 100$
3	PM_STD	$std(PM_i)$
4	AM_STD	$std(AM_i)$
5	STD of PM - AM	$std(PM_i - AM_i)$
6	STD of PM-to-next-AM	$std(AM_{i+1} - PM_i)$
7	PM-to-next-PM correlation	$E\left[\Delta T_{2,i}\cdot\Delta T_{2,i+1}\right]$
8	PM-to-next-AM correlation	$E\left[\Delta T_{2,i}\cdot\Delta T_{1,i+1}\right]$
9	AM-to-next-AM correlation	$E\left[\Delta T_{1,i}\cdot\Delta T_{1,i+1}\right]$
10	AM-to-PM correlation	$E\left[\Delta T_{1,i}\cdot\Delta T_{2,i}\right]$
11	PM-to-next-PM correlation coefficient	$\frac{E\left[\Delta T_{2,i}\cdot\Delta T_{2,i+1}\right]}{\sqrt{E\left[\Delta T_{2,i}\right]^2\times E\left[\Delta T_{2,i+1}\right]^2}}$
12	PM-to-next-AM correlation coefficient	$\frac{E\left[\Delta T_{2,i}\cdot\Delta T_{1,i+1}\right]}{\sqrt{E\left[\Delta T_{2,i}\right]^2\times E\left[\Delta T_{1,i+1}\right]^2}}$
13	AM-to-next-AM correlation coefficient	$\frac{E\left[\Delta T_{1,i}\cdot\Delta T_{1,i+1}\right]}{\sqrt{E\left[\Delta T_{1,i}\right]^2\times E\left[\Delta T_{1,i+1}\right]^2}}$
14	AM-to-PM correlation coefficient	$\frac{E\left[\Delta T_{1,i}\cdot\Delta T_{2,i}\right]}{\sqrt{E\left[\Delta T_{1,i}\right]^2\times E\left[\Delta T_{2,i1}\right]^2}}$

3.3.2 Parameter Regularization

Due to the small number of samples, we adopted leave-one-out cross validation (jackknife) for parameter regularization of the classifiers. For each of the (2^7-1) possible combinations of input features, we trained the SVM classifier, choosing as the optimal kernel size σ and cost parameter C for the SVM (from discrete sets) that gave the best validation performance. Validation performance is the average of the classification rate of 35 classifiers, each of which is trained on a different 34-sample training set, and evaluated on a one-sample validation set. The discriminative ability of each feature combination is measured using the validation accuracy achieved with the optimal parameters.

4. GENERALIZATION PERFORMANCE

Having selected the optimal feature set, we want to estimate the classifier performance on *an independent test set* not used for training or adjusting regularization. Again, the scarcity of samples suggests we adopt leave-one-out cross-validation scheme. With a regularization parameter to choose, this is a two-loop cross validation-process (O(N(N-1)) complexity in the number of samples). We sequentially select one subject as the *test set*, and use the remaining 34 subjects as the *development set*. In the development set, we do another leave-one-out cross-validation to determine the optimal kernel size σ and cost parameter C for the SVM. Then we train a classifier on the complete development set and apply it to the single-sample test set. We repeat this over all 35 development-test partitions and report the total number of misclassified test samples as the test set error.

We conducted this exhaustive validation of SVM performance using features {1, 2, 3, 4, 10, 12, 13}. The feature combinations {1, 2, 3}, {1, 2, 12}, and {1, 2, 13} achieved the best validation results which equaled to 0.83. Therefore we conducted the tests on these three feature configurations. The test results are shown in Table 2. The feature set {1, 2, 3} misclassified 9 of the 35 examples (correct rate 0.74). The 95% confidence limit computed using binomial statistics assuming a Bernoulli trial model on this classification rate is [0.58, 0.86].

5. DISCUSSION

The results show that the viability of making reliable inference for mild cognitive loss on the individuals from the MedTracker data. It is hard to expect good performance due to the small number of samples and the short data streams. Since not all of the features are equally useful for discriminating between the LCP and HCP groups, selecting proper subset of candidate features is crucial. Future work will investigate the application of statistical pattern recognition techniques to long term medication adherence for the purpose of detecting emerging cognitive decline and identifying change points.

Table 2: SVM Test on 35 subjects (validation result = 0.83)

Feature Combination	Accuracy
1+2+3	**0.74**
1+2+12	0.60
1+2+13	0.63

ACKNOWLEDGEMENTS

The authors thank Tamara Hayes and Jeffrey Kaye for technical discussion and Tian Lan for help with feature ranking using mutual information. Data was collected by OHSU ORCATECH under NIA grant P30 AG024978. This work was supported by Intel under the OHSU BAIC project and partially by NSF grants ECS-0524835, ECS-0622239, and IIS-0713690.

REFERENCES

[1] J.J. Branin, "The role of memory strategies in medication adherence among the elderly," *The Journal of Home Health Care Services Quarterly*, vol. 20(2) pp.1-16, 2001.

[2] T.L. Hayes, J.A. Kaye, M. Pavel, J. Lundell, and J. Kimel, "Effect of early cognitive changes on medication adherence," in *59th Annual Scientific Meeting of the Gerontological Society of America*, Dallas, TX, 2006.

[3] T.L. Hayes, J.M. Hunt, A. Adami, and J.A. Kaye, "An electronic pillbox for continuous monitoring of medication adherence," in *28th Annual Int. Conf. of IEEE EMBS,* pp.6400-6403, 2006.

[4] D.P. Graham, J.A. Cully, A.L. Snow, P. Massman, and R. Doody, "The Alzheimer's disease assessment scale - cognitive subscale: normative data for older adult controls," *Alzheimer Disease and Associated Disorders*, vol.18, pp.236-240, 2004.

[5] T. Leen, Z. Lu, T. Hayes and J. Kaye, "Detection of early cognitive loss from medication adherence behavior," in *2nd International Conference on Technology and Aging,* Toronto, Canada, 2007.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd Edition, Boston, Academic Press Inc., 1990.

[7] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification,* 2nd Edition, New York, John Wiley & Sons, Inc., 2001.

[8] C. J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery,* vol. 2, pp.121-167, 1998.

[9] T. Lan, D. Erdogmus, "Maximally informative feature and sensor selection in pattern recognition using local and global independent component analysis," *Journal of VLSI Signal Processing Systems*, to appear in 2007.