Detecting Parkinsons' Symptoms in Uncontrolled Home Environments: A Multiple Instance Learning Approach

Samarjit Das, Breogan Amoedo, Fernando De la Torre and Jessica Hodgins

The Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA

Abstract— In this paper, we propose to use a weakly supervised machine learning framework for automatic detection of Parkinson's Disease motor symptoms in daily living environments. Our primary goal is to develop a monitoring system capable of being used outside of controlled laboratory settings. Such a system would enable us to track medication cycles at home and provide valuable clinical feedback. Most of the relevant prior works involve supervised learning frameworks (e.g., Support Vector Machines). However, in-home monitoring provides only coarse ground truth information about symptom occurrences, making it very hard to adapt and train supervised learning classifiers for symptom detection. We address this challenge by formulating symptom detection under incomplete ground truth information as a multiple instance learning (MIL) problem. MIL is a weakly supervised learning framework that does not require exact instances of symptom occurrences for training; rather, it learns from approximate time intervals within which a symptom might or might not have occurred on a given day. Once trained, the MIL detector was able to spot symptom-prone time windows on other days and approximately localize the symptom instances. We monitored two Parkinson's disease (PD) patients, each for four days with a set of five triaxial accelerometers and utilized a MIL algorithm based on axis parallel rectangle (APR) fitting in the feature space. We were able to detect subject specific symptoms (e.g. dyskinesia) that conformed with a daily log maintained by the patients.

Index Terms— Parkinson's Disease (PD), continuous motor symptom monitoring, multiple instance learning

I. INTRODUCTION

Recently, there has been an increased interest in developing wearable sensing systems for patient-centric healthcare. The ultimate goal of these collaborative efforts between the healthcare and engineering communities is to enable unobtrusive autonomous monitoring of the patients' state and generate valuable clinical feedback. In that regard, motor symptom monitoring in Parkinson's Disease (PD) has gained significant attention over the years [11], [15]. In most cases, these symptoms (e.g. tremors, dyskinesia) fluctuate during daily living depending on the medication intake. The knowledge of these medication cycles could be very useful for PD treatments including: a) evaluation of the potential benefit of deep brain stimulation (DBS), b) patient adapted drug therapy and c) tracking disease progression over time. However, due to the limited frequency and duration of clinic visits, robust assessments of medication cycles is difficult and often relies on self-reports by the patients. Hence, there is a need for sensing systems capable of unobtrusive continuous monitoring of the motor symptoms in uncontrolled daily living environments.
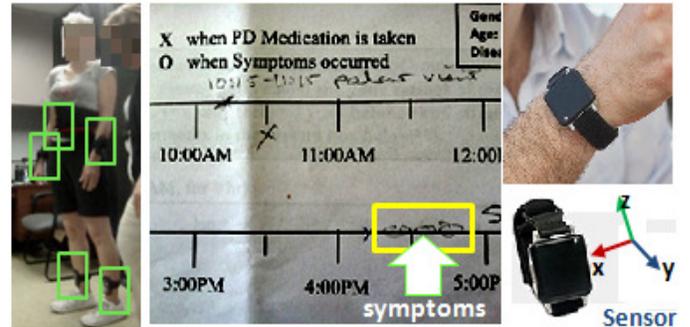
Fig. 1. The sensing system and a daily log that the patients maintained. The green boxes indicate accelerometer locations. In the log, the symbol *X* indicates medication intake and *O* indicates symptom occurrences (indicated by the arrow). The motion sensors (APDM Inc.) are shown on the far right.

Several researchers have addressed the problem of quantitative assessment of PD motor symptoms [12], [3], [11], [6]. In [11], researchers explored the feasibility of using accelerometer data to estimate the severity of motor symptoms symptoms using a support vector machine (SVM) classifier [4]. In [19], the authors developed a monitoring system that provided motor assessments by simultaneously analyzing current motor activity of the patients (e.g., sitting, walking) in a simulated daily living environment. Other related works performing accelerometer-based PD motor symptom assessment include [7], [17], [9].

Most of the existing work involves supervised learning frameworks under controlled laboratory settings or simulated home environments. The key difficulty in monitoring within uncontrolled environments is the lack of reliable ground truth information, which makes the use of supervised learning frameworks very challenging. The inherent problem lies in the fact that accurate labeling of human behavior is error prone, subjective and time consuming. These issues are accentuated when self-report is the only form of measurement available. We address these by using a weakly supervised learning framework known as multiple instance learning (MIL) [5]. Weakly supervised learning has found many applications in recent years including objection detection in computer vision [10]. MIL does not require accurate labels for every time instant. While training, it only needs to know whether or not a time interval contains discriminative patterns (e.g. a symptom) *anywhere* within it. MIL automatically learns those patterns and localizes them during the test phase (i.e. detect time intervals containing symptom occurrences as well as identify symptom instances).

We monitored two PD patients, each for four days with a

set of five triaxial accelerometers (located at the waist and limbs). Our detector was able to spot the subject specific motor symptoms that conformed with the daily log maintained by the patients.

II. THE MONITORING SYSTEM

Our home monitoring system comprised of five wearable triaxial accelerometers sampling at 40 Hz with 16 bit resolution. Each unit was capable of continuous data logging for about 20 hours and their batteries were recharged at night. The devices were strapped to the wrists, ankles and waist as shown in Fig. 1. The subjects were asked to wear them during all waking hours of the day (except water activities) and to complete their regular daily living activities as normal. There was no predefined scripted activity performance during the monitoring period. The subjects were asked to maintain a daily log of their medication intakes as well as to note the approximate time intervals of symptom occurrences on a time scale (e.g. within a range of 20 minutes to half an hour). A snapshot of the daily log is shown in Fig. 1.

III. FEATURE EXTRACTION

Prior work on accelerometer based activity recognition [2] utilized various features like mean, energy, high frequency energy content, correlation and frequency domain entropy extracted from multiple accelerometers. While we use similar features, we put particular emphasis on the choice of features depending specific PD symptoms.

We extracted several features from the raw accelerometer data over six second windows with one second overlap. Similar window sizes have been found to be optimal for accelerometer based activity analysis [2], [11]. Features included mean, energy, high frequency energy content, correlation and frequency domain entropy as well as a five bin histogram representation of the spectral contents over all three axes (similar to a bag of words representation [13] in the frequency domain). The feature extraction was particularly targeted at two specific motor symptoms that were prominent in our subject pool: a) dyskinesia in the left hand and b) a shaky left hand with predominant left-right asymmetry. For dyskinesia detection, we chose frequency domain histogram, high frequency energy content (beyond 4Hz) and frequency domain entropy, all extracted from the accelerometer in the left wrist. Similar features were also found to suit well for detecting left-right asymmetry by using discriminative measures across the two wrists (e.g., difference in high frequency contents). These features, adapted to individuals and their specific symptoms, are essential for building person-specific models for home monitoring.

IV. MULTIPLE INSTANCE LEARNING

Multiple Instance Learning (MIL) [5], [1], [10] is proposed as a weakly supervised learning algorithm for problems with incomplete knowledge about the labels of training examples. In supervised learning, every training instance (e.g., a feature vector from the accelerometer data) is assigned with a discrete or real-valued label (e.g. dyskinesia or normal activity).
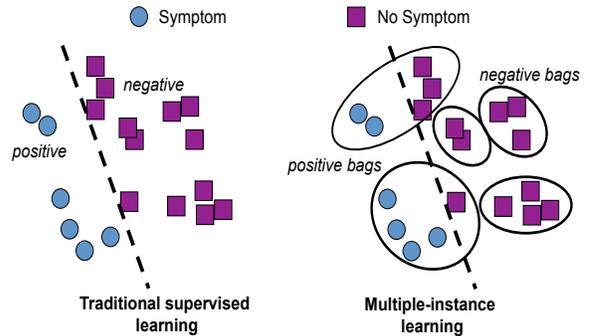

Fig. 2. The intuitive idea behind multiple instance learning (MIL) in contrast to supervised learning approaches.

In comparison, for MIL, the labels are only assigned to bags of instances (i.e. a collection of feature vectors over a time interval). A bag is labeled as positive if at least one instance in that bag is positive, and the bag is labeled negative if all the instances in it are negative (see Fig. 2). In other words, if a symptom occurred at *any* time within the interval, then the entire interval receives the corresponding symptom label and it is labeled symptom-free *only if* no instance within that interval is associated with a symptom. Thus, for training, it is not necessary to know the exact instances of symptom occurrences i.e. there is no label for individual instances. The goal of MIL is to classify unseen bags of instances based on the labeled bags as training data. Some variants of MIL [5], [10] also estimate the instance labels during the test phase, which would imply temporal localization of the symptom instances within the time segments.

The problem definition of MIL can be posed as follows. Consider a set of feature vectors $\mathbf{x}_i$, $i = 1, 2, \ldots, N$. Under a supervised learning framework, each feature vector i.e. instance would have a labels $l_i = f(\mathbf{x}_i) \in \{+1, -1\}$. In contrast, a weakly supervised MIL framework does not have labels for each instance, rather, it has labels for a set of instances (a bag). Consider a feature space with B bags. The bth bag with m_b feature vectors $\mathbf{X}_b = [\mathbf{x}_1^b, \mathbf{x}_2^b, \ldots, \mathbf{x}_{m_b}^b]$ will be labeled $L_b = f(\mathbf{X}_b) \in \{+1, -1\}$ as follows

$$f(\mathbf{X}_b) = \begin{cases} 1 & \text{if} \quad \exists j \in \{1, \ldots m_b\} \text{ with } g(x_j^b) = 1 \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

where $b = 1, 2, \ldots B$. In other words, if the observed bag label is positive then at least one of the variant instances must have produced that positive result. Furthermore, if the observed label is negative, then none of the variant instances could have produced a positive result. This rule can be modeled by introducing the function $g(\mathbf{x}_i^b)$ that takes a single instance and produces a result. During training, we will only know the external bag labels i.e. $L_b = f(\mathbf{X}_b)$ and do not know the internal function $g(\cdot)$. The ideal goal of the MIL algorithm is to construct an approximation $\hat{g}(\cdot)$ of $g(\cdot)$. A hypothesis $\hat{g}$ is consistent with a set of training examples if it classifies every feature vector of every negative example as negative while classifying at least one feature vector of every positive example as positive. Thus, under this framework, given a test bag of instances $\mathbf{X}_b^{test}$, we would be able to predict the

corresponding label $f(\mathbf{X}_b^{test}) \in \{+1, -1\}$ as well as compute $\hat{g}(\mathbf{x}_i^b)$, $\mathbf{x}_i^b \in \mathbf{X}_b^{test}$. Some variants of MIL only predict bag labels $f(\mathbf{X}_b^{test})$ during the test phase. The main difference of the MIL framework from supervised learning is that each training example is represented by multiple instances (hence the name).

A. MIL Algorithms

The solution to the MIL problem was first proposed in [5]. It involved finding an axis-parallel hyper-rectangle (APR) in the feature space that captures the target concept. Intuitively, this APR should contain at least one instance from each positive bag and meanwhile exclude all the instances from negative bags. In [5], the authors suggested three variants of the APR algorithm: 1) a "standard" algorithm that finds the smallest APR which bounds all the instances from positive bags, 2) an "outside-in" algorithm for constructing the smallest APR that bounds all the instances in positive bags and then shrinks the APR to exclude false positives and, 3) an "inside-out" algorithm that starts from a seed point and then grows a rectangle from it with the goal of finding the smallest APR which covers at least one instance per positive bag and no instances from negative bags. For details please refer to [5].

In our current work, we use an iterative, discriminative variant of the APR algorithm [5] (ID-APR) for multiple instance learning. It is a combination of the standard and outside-in APR algorithms. It treats the time series monitoring data in terms of temporal windows of fixed lengths where each instance corresponds to a single window. We compared the performance of ID-APR with several other MIL algorithms that use a similar formulation (more details in the next section). Algorithms like detection-segmentation SVMs [10] perform weakly supervised learning on time series data and pull out discriminative *noncontiguous* time segments across two time series classes. However, these methods involve optimizations that tend to have many local minimas and this issue poses a major hurdle for training with large datasets that include uncontrolled activities.

V. EXPERIMENTS AND RESULTS

In our preliminary study, we had two PD patients (one male and one female) participate. The male subject was aged 55 years with a disease duration of 6.5 years whereas the female was aged 54 years with a disease duration of 6 years. Each individual wore our monitoring system continuously except during sleep for four days and continued their regular daily activities. They maintained a daily log as mentioned in Sec. II. Our study protocol was approved by Carnegie Mellon University's Institutional Review Board (IRB) for human subject protection. Each patient was interviewed beforehand in order to gather information about their disease state as well as their specific motor complications. This information guided us during the feature selection process. The female patient took her PD medication four times a day and had predominant left hand dyskinesia, mostly around medication intake times. The male patient took his PD medications three

times a day and suffered from a shaky left hand, mostly during the morning hours.

The patient logs provided us with approximate time intervals which might contain symptom occurrences. We trained an APR based [5] multiple instance learning algorithm with chunks of accelerometer data of 20-40 minute durations (i.e. bags containing hundreds of instances or features vectors). We assigned the bags with positive or negative levels depending on whether or not it contained a symptom occurrence. Our monitoring system was implemented using the ID-APR algorithm of MIL Toolkit [16]. The training was performed on eight hours of monitoring data collected during day 1. Then, provided with the test data from the same day or a different day, the algorithm was able to spot 20-40 minute segments where symptoms occurred and provide an estimate of when it happened within the segment. In Fig. 3, we demonstrate the results over a six hour monitoring period for the female patient with dyskinesia. The ground truth information acquired from the patient logs is compared against the output of the MIL detector and the results are promising. Similar experiments were performed for detecting the symptoms of the male patient and the results are shown in Fig. 4. The predominance of his motor symptoms during the morning hours can clearly be seen as picked up by the MIL detector.

Next, we compared the performance characteristics of the ID-APR based MIL detector with other MIL algorithms like multiple instance SVM (MI-SVM) [1], citation k-nearest neighbor (kNN) [14], diverse density (DD) [8], expectation maximization version of DD [18]. For each patient, we tested hours worth of monitoring data over 50 time segments extracted from all the four days (each with lengths greater than 20 minutes containing symptom or no symptom labels). Then we analyzed the performance characteristics of the algorithms by computing their bag level accuracies (i.e. percentage of total bags correctly classified as symptom or normal). As it can be seen in Fig. 5, the ID-APR based MIL algorithm outperforms the rest of the algorithms. These preliminary results indicate that multiple instance leaning
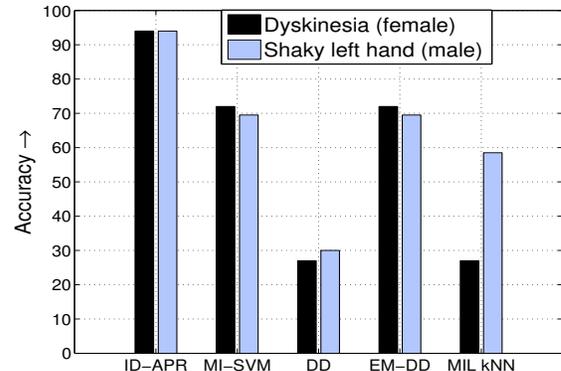

Fig. 5. The performance comparison of ID-APR algorithm with other MIL algorithms like multiple instance SVM (MI-SVM) [1], citation k-nearest neighbor (kNN) [14], diverse density (DD) [8], expectation maximization version of DD [18]. As it can be seen that ID-APR has the highest accuracy in detecting the PD symptoms.
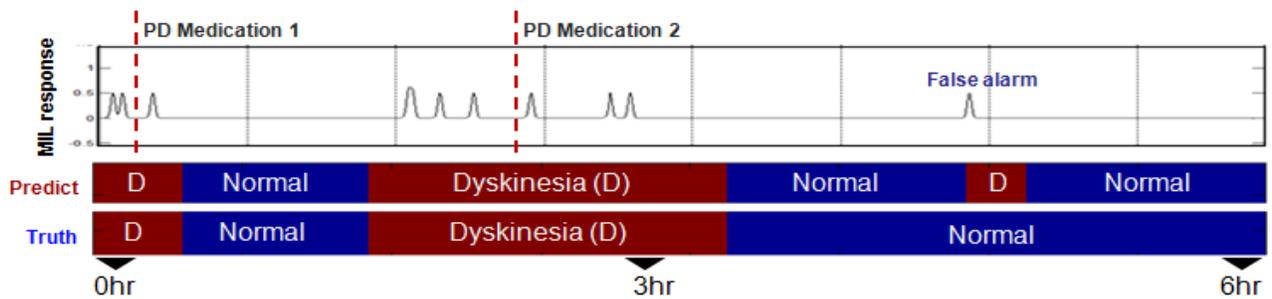
Fig. 3. Detection of dyskinesia for the female patient during a six hour monitoring period. The spikes indicates the estimated instances of symptom occurrences as predicted by the MIL detector. Notice the prevalence of motor symptoms around the medication times.
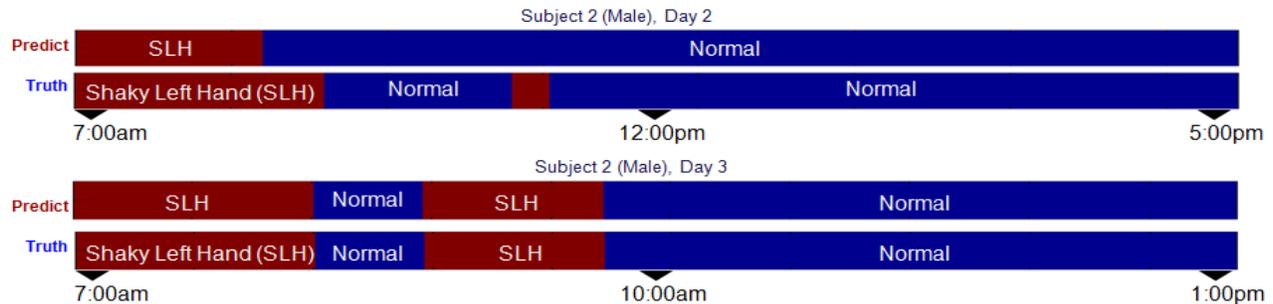

Fig. 4. Monitoring of the motor symptoms of the male subject over several hours across two days. The prominence of his symptoms (shaky left hand) during the morning hours is clearly visible as predicted by the MIL detector.

could potentially be a promising direction for developing monitoring systems capable of learning symptom characteristics from partial ground truth information and detecting them in an uncontrolled daily living environments.

VI. CONCLUSION

In this paper, we have demonstrated the use of multiple instance learning for detecting PD motor symptoms in uncontrolled home environments. Our work addressed the formulation of PD symptom detection from weakly labeled data as a semi-supervised multiple instance learning problem. The features were carefully chosen to address the subject and symptom specific nature of the problem. We show promising preliminary results on four days of monitoring performed with two PD subjects. In future work, we plan to increase our subject pool and utilize optimal feature selection strategies under MIL frameworks for developing robust person-specific models. These techniques can potentially be adapted to various other physiological sensing and monitoring applications as well.

REFERENCES

[1] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems 15*, pages 561–568. MIT Press, 2003.

[2] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Intl Conference on Pervasive Computing*, 2004.

[3] P. Bonato, D. M. Sherrill, D. G. Standaert, S. S. Salles, and M. Akay. Data mining techniques to detect motor fluctuations in Parkinsons disease. In *Proc. Conf. IEEE Eng. Med. Biol. Soc*, 2004.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 1995.

[5] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.*, 89(1-2):31–71, Jan. 1997.

[6] N. L. Keijsers, M. W. Horstink, and S. C. Gielen. Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Movement Disorder*, 18:70–80, 2003.

[7] N. L. Keijsers, M. W. Horstink, and S. C. Gielen. Ambulatory motor assessment in Parkinsons disease. *Movement Disorder*, 2006.

[8] O. Maron and T. Lozano-Prez. A framework for multiple-instance learning. In *Conf. on Advances in Neural Information Processing Systems*, 1998.

[9] S. Moore, H. MacDougall, J. Gracies, H. Cohen, and W. Ondo. Long-term monitoring of gait in Parkinson's disease. *Gait Posture*, 2007.

[10] M. H. Nguyen, L. Torresani, F. De la Torre, and C. Rother. Weakly supervised discriminative localization and classification: a joint learning process. In *Proceedings of ICCV*, 2009.

[11] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato. Monitoring motor fluctuations in patients with Parkinsons disease using wearable sensors. *IEEE Trans. on Information Technology in Biomedicine*, 2009.

[12] M. Pavel, T. Hayes, A. Tsay, W. Erdogmus, A. Paul, N. Larimer, H. Jimison, and J. Nutt. Continuous assessment of gait velocity in Parkinsons disease from unobtrusive measurements. In *EMBC*, 2007.

[13] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering objects and their location in images. In *Proceedings of CVPR*, 2005.

[14] J. Wang and J. D. Zucker. Solving the multiple-instance problem: a lazy learning approach. In *Proc. 17th Int'l Conf. on Machine Learning*, 2000.

[15] W. J. Weiner. Motor fluctuations in Parkinsons disease. *Rev. Neurol. Dis.*, pages 101–108, 2006.

[16] J. Yang. A multiple instance learning library (CMU, 2008). URL: http://www.cs.cmu.edu/ juny/mill).

[17] H. Zabaleta, T. Keller, and J. M. Masso. Gait analysis in frequency domain for freezing detection in patients with Parkinson's disease. In *ISG*, 2008.

[18] Q. Zhang and S. A. Goldman. Em-dd: An improved multiple-instance learning technique. In *In Neural Information Processing Systems*, 2001.

[19] D. G. Zwartjes, T. Heida, J. P. V. van, J. A. Geelen, and P. H. Veltink. Ambulatory monitoring of activities and motor symptoms in Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 57(11), 2010.