Identification of Parkinsonian gait by means of inertial measurements units via metric learning algorithms

Dr Fabio Cuzzolin¹, Mr Michael Sapienza¹, Dr Patrick Esser², Prof Helen Dawes², Miss Marloes Franssen³

  1) Artificial Intelligence and Vision Group, Department of Computing and Communication Technologies, Oxford Brookes University, Oxford, UK
  2) Movement Science Group, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK

Contact details:
Dr Fabio Cuzzolin
Artificial Intelligence and Vision Group
Department of Computing and Communication Technologies
Oxford Brookes University, Wheatley campus
Oxford, OX33 1HX
United Kingdom
Tel: +44 (0) 1865 484526
@: Fabio.cuzzolin(at)brookes.ac.uk

Key Words:
Machine Learning Algorithms, Hidden Markov Models, Metric Learning, Inertial Measurement Unit, Gait, Parkinson’s

Notes:
Journal Experimental Biology, submission date February 2014
Summary (<150)
Introduction

Parkinson’s disease is the second most common neurodegenerative disorder with a prevalence of 1-4% above the age of 60\(^1\). PD presents itself with both motor and non-motor symptoms. The clinical diagnosis of Parkinson’s disease (PD) according to the UK Brain Bank criteria is based on the presence of motor symptoms (such as bradykinesia, rigidity and tremor) and response to dopaminergic medication\(^6\). As such, the monitoring of these motor symptoms can offer a means to detect and monitor disease progression. In established PD the Brain Bank criteria applied by experts shows 90% sensitivity and specificity for the presence of midbrain Lewy bodies\(^4\). However, in early disease clinical diagnosis is less straightforward and PD diagnosis made in the community by non-experts is still associated with a 25% error rate\(^4\). As such there is a need for a more automated diagnostic and monitoring tool to support diagnosis in primary care\(^4\).

Walking has been indicated as a sensitive indicator for the progression of Parkinson’s\(^5\) as individuals present an altered gait pattern with increased cadence and reduced stride lengths\(^6\).

With advances in technology, measuring gait in larger cohorts is becoming more accessible to those interested. Inertial Measurement Units (IMU) can be used to obtain objective measurements of gait parameters inexpensively, quickly and easily in clinical environments\(^6\). These methods have been validated within clinical conditions such as Parkinson’s\(^5\). However, simple temporal (steptime and cadence) or spatial (stride length and walking speed) parameters are not sensitive indicators of early Parkinsonian gait over short walking distances in their own right (i.e., when classified directly). Whereas, recent analysis of Centre of Mass (CoM) excursion variations via phase plot analysis has indicated the potential of using more sophisticated classification methodologies for detection of Huntington’s (Collett, Esser, in press) and Parkinson’s\(^6\).

The automatic classification of human disease condition from IMU data is difficult as human gait possesses a high degree of variability between subjects of the same class. For instance a person’s movements may vary in temporal or spatial parameters, yet still belong to the same disease category. Moreover, one needs to generalise over several nuisance factors such as the variation of disease progression, age of the subject and the electronic noise present in the measurement signals whilst still being able to discriminate between various disease categories.

Whilst it has been suggested that machine learning algorithms (MLA) can be used for classification of Parkinson’s by means of gait measurements\(^9\)\(^10\), these results have only been investigated in small numbers (27\(^9\), 20\(^10\)) via CoM-IMU collected data. In contrast, our work includes an increased sample (5-fold) of clinical data, covering a wide range of severities of PD in addition to a large cohort (n=2168) of typical developed adults. In addition, whereas the studies by Barth et al.\(^9\) and Cancela et al.\(^10\) used predefined, fixed-length features and standard off-the-shelf classifiers, we propose to learn dynamical models from the time-series gait measurements, and use a tailored classification method.

Recent work in which an optimal classifier for the problem at hand is constructed from the training data via metric learning techniques (cite Xing, Lebanon) has shown promising results in classifying human action image sequences represented as a time series of histograms\(^11\).
This study therefore sets out to explore whether the novel application of these optimal metric learning based classifier can: firstly, automatically distinguish those with and without Parkinson’s, during a clinically accepted 10-metre walk test, within a large cohort; and secondly, automatically determine disease severity.
Materials and Methods

Experimental setup

Subjects

Previous recoded gait data on 2168 typical developed adults (TDA), ranging from 5-80 years, at the Science Museum London was reanalysed. Furthermore, gait data of 163 people with Parkinson’s (PwP) were reanalysed. Each participant’s date of birth and leg length was recorded. PwP were additionally assessed on disease severity using the Movement Disorder Society - Unified Parkinson’s Disease Rating Scale (MDS-UPDRS), ranging from 0 (highest motor disability) to 100 (indicating no motor symptoms). In addition, from the health status short-form health survey (SF-36) walking components were extracted and used for classification of motor symptoms, ranging from 0 to 100 where a higher score represents a more favourable health state.

Protocol

Gait measurements were performed by means of an inertial measurement unit (IMU) which was attached to the lower spine (Lumbar 4 region) over the projected CoM by means of double sided adhesive tape. Participants in both studies were instructed to walk over a 10 metre walkway free of obstacles at their self-selected walking speed.

Classification in Machine Learning

In order to automatically determine whether a person has Parkinson’s disease and its severity from the IMU data we turned to the field of Machine Learning and data mining.

From a Machine Learning point of view, the problem is defined as follows. Given a repository of \( n \) gait motions \( G_k \) (a “training set”) \( D = \{ (G_1, Y_1), \ldots, (G_n, Y_n) \} \), to each of which is assigned a “class label” \( Y_k \) (e.g. normal versus Parkinson), we want to learn an appropriate machinery (a “classifier”) which, given as input a new, unlabelled gait motion, produces the class label of the new sequence.

In the following, training gait motions can assume two types of labels:

1) Binary, normal \( 0 \) versus Parkinson \( 1 \);

2) Nominal severity score in the range 100 to 20, where lower scores indicate higher severity.

We therefore have two classification problems: the first one in which it is decided whether the subject performing the gait motion is affected by Parkinson or not; the second one whose output is instead the degree of severity of such a condition.

Solving a classification problem involves a number of steps, including:

i) Finding the more suitable representation for the input data;

ii) Designing/selecting the most appropriate classifier for the problem at hand.
In our experimental setting, each instance of gait motion is represented by a time series of inertial measurements (IMU). However, IMU sequences may be of different lengths, so we need to find a constant-size representation for all gait sequences (time warping\textsuperscript{13}). In addition, studies in gesture and gait classification indicate that the dynamics of time series of measurements can be very discriminative, i.e., can much help with their classification\[REF\].

Researchers have explored the idea of encoding motions via linear, nonlinear\textsuperscript{14} or chaotic\textsuperscript{15} dynamical systems. Hidden Markov models\textsuperscript{16}, in particular, have been widely employed\textsuperscript{13}, as they provide a solution to the time warping problem while efficiently encoding the dynamics of the motion to classify.

**Hidden Markov Models for representing time series**

A hidden Markov model (HMM) is a finite-state stochastic model whose states form a Markov chain. Transitions between states are governed by a transition matrix $A$, which specifies the probability of passing from state $i$ to state $j$, for each pair of states (Fig. 1).

![Fig. 1. A Hidden Markov model is a finite state automata, described by the matrix $A$ of transition probabilities between each pair of states, and a Gaussian output probability for each state which determines the likelihood $p(y|x=i)$ of generating an observation $y$ given state $i$.](image)

Although the state of an HMM is hidden (i.e., it cannot be observed directly), it generates in a linear way an observation (in general, a vector of observations) which can instead be observed. For each state $i$, an output Gaussian distribution $\Gamma_i$ tells us the likelihood with which an observation $y$ is generated by state $i$.

Given a sequence of measurements $(y_1, ..., y_T)$ (in our case, a sequence of IMU associated with a walking gait), the HMM most likely to generate it can be identified via the Expectation Maximisation (EM) algorithm\textsuperscript{16}. The dynamics of each walking gait, regardless its length, can then be represented via a HMM $H=(A,C)$ with the same number of states (a parameter of the EM algorithm, which needs to be tuned to the data at hand), where $A$ is the transition matrix and $C$ is the matrix collecting the means of the Gaussian output densities.

Classifying walking gaits (in terms of either presence of a condition, or degree of severity) reduces in this way to classifying the hidden Markov models learnt from the series of IMU measurements.
Classifying times series via Hidden Markov Models

HMMs are typically classified by learning a new model for each test sequence, measuring its distance from the old models, and attributing to it the label of the closest model(s).

A number of distance functions for dynamical systems (e.g. Cauchy kernels\textsuperscript{14}) and HMMs in particular\textsuperscript{17} have been introduced. Nevertheless, no single distance function can possibly outperform all the others in every classification problem, as the same models (or image sequences) can be endowed with different labels (e.g., as in our case, presence of a disease or disease severity).

A sensible approach when training data are available consists in learning in a supervised fashion the “best” distance function for a specific classification problem, e.g. by maximising the classification performance achieved.

This approach is widely supported by the literature, particularly in the case of distances induced by a linear transformation of the original data domain\textsuperscript{18}. However, dynamical models (as HMMs in particular) live in highly nonlinear spaces. The need for a principled way of learning distances in general metric spaces arises.

Learning optimal metrics for classification

Recent work has shown promising results classifying human action sequences represented as a time series of histograms\textsuperscript{11}. The inertial measurements represent a multidimensional time-series of variable length per person and thus the same ideas presented in \textsuperscript{11} may be employed here.

The main idea is to initially encode the IMU data via dynamical models such as the hidden Markov model (HMM), which have shown to be effective at representing sequences of variable length and proven to be effective in both action recognition and gait identification\textsuperscript{11}.

In this case, disease recognition reduces to classifying dynamical models by learning an IMU-HMM for the unknown test subject, and measuring its distance to the trained models. An outline of the disease recognition pipeline is illustrated in Fig. 2.

![Fig. 2: Overview of Machine learning algorithm proposed by\textsuperscript{11} for the classification of time-series data.](image-url)
Data analysis

An experiment was setup to determine how much better our optimal metric learning classifier was at predicting the disease labels for both control and Parkinson’s subjects as compared to a machine randomly assigning a label to each test subject (random guessing). Moreover, the disease’s degree of severity was also estimated.

Gait representation

Gait parameters were derived from excursion of the CoM measured by IMU6. Spatial (stride length and walking speed) temporal (step time and cadence) parameters were derived according to well described and validated inverted pendulum models6.

Classification of PwP versus TDA

For each sequence of IMU measurements the parameters of an HMM with n=3 states were learned via the Expectation Maximisation (EM) algorithm19. Since EM suffers from local minima, the algorithm was applied 10 times for each sequence and the parameters yielding the highest likelihood were kept. The EM convergence threshold was set to $10^{-5}$, and limited to 50 iterations if convergence was not reached.

We learned an optimal pullback distance by maximising the classification performance on the training set. As base distance between two HMMs $H_1=\{A_1,C_1\}$ and $H_2=\{A_2,C_2\}$ we used the Frobenius norm $|A_1-A_2|_F+|C_1-C_2|_F$, where $|M|_F=\sqrt{tr(M^T M)}$. We used with 5-fold cross-validation, i.e., we divided the training set into 5 folds and we used a fold of testing and the other 4 for training, for all choices of the testing fold, and measured the average performance. Classification was performed via Nearest Neighbour (1-NN): each test HMM was attributed to the class of the nearest model in the training set according to the learned optimal distance measure. The disease severity for PwP was estimated by averaging the severity levels from the 5 nearest neighbours of the predicted sample (Fig. 3).

Fig. 3: Disease severity for a test IMU sequence, represented as a HMM, is estimated by locating its 5 nearest neighbours (according to the optimal distance learned) and taking the average of their severity levels.

In order to estimate how well the MLA algorithm will perform in practice, we used repeated random sub-sampling validation. We applied our classification algorithm 25 times, randomly sampling new train and test sets at each run. For each run, ⅔ of the data from each disease category was randomly sampled as training data, and the remaining ⅓ was held for validation.

Results

Subject demographics
Participant descriptive data was expressed as mean±standard deviations (range min-max). For PwP, the MDS-UPDRS motor subsection (part 3) and the SF-36 physical functioning sub-score were assessed by calculating the median (range min-max).

In total 160 PwP were measured over a 10metre walkway and included in this study. Participants were found to be 67.2±8.0 years of age. IMU derived walking speed for PwP was found to be 1.12±0.18 ms\(^{-1}\) (range 0.59-1.70 ms\(^{-1}\)). PwP were assessed by the MDS-UPDRS on which they scored 17 (range 0-57) on the motor section part 3. Furthermore participants were found to score 75 (range 20-100) on the physical functioning section of the SF-36. The total SF-36 score was found to be 71 (range 25-94) for PwP.

Additionally 2168 typical developed adults (TDA) were included in this study whom were found to be 26.8±15.5 years (range 5-80 years) of age. On average their IMU derived walking speed was found to be 1.39±0.19 ms\(^{-1}\) (range 0.85-1.98 ms\(^{-1}\)).

Results of disease classification and severity prediction

The classification results were initially expressed as a confusion matrix, which summarises the predicted and actual classes of the test samples estimated by the classifier, as shown in Table 1. In this case, a false positive (fp) occurs when a person is predicted with Parkinson’s but does not actually have the disease. A false negative (fn) occurs when a person is predicted to be a TDA when they actually have Parkinson’s. Both (tp) and (tn) denote the true positive and true negative values respectively.

The true positive rate (TPR, also called sensitivity or recall) is calculated as: TPR = tp/(tp+fn). The precision is calculated as: Precision = tp/(tp+fp). The accuracy is defined as: ACC = (tp+tn)/(tp+tn+fp+fn), and the F1 score, which is the harmonic mean of the precision and sensitivity is: F1 = 2*tp/(2*tp +fp+fn).

<table>
<thead>
<tr>
<th>MLA</th>
<th>Pred.</th>
<th>PwP</th>
<th>TDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gt.</td>
<td>PwP</td>
<td>tp=27.9</td>
<td>fn=25.1</td>
</tr>
<tr>
<td></td>
<td>TDA</td>
<td>fp=14.3</td>
<td>tn=708.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random</th>
<th>Pred.</th>
<th>PwP</th>
<th>TDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gt.</td>
<td>PwP</td>
<td>tp=26.9</td>
<td>fn=26.1</td>
</tr>
<tr>
<td></td>
<td>TDA</td>
<td>fp=360.4</td>
<td>tn=362.6</td>
</tr>
</tbody>
</table>

The accuracy of the proposed classification approach in determining PwP from TDA, averaged over the 25 repeated runs, was (94.92±1.01%), compared to (50.19±1.76%) when classified according to random chance (50% accuracy is expected for random guessing on a binary classification problem). Due to the imbalance of the number of TDA and PwP samples, the F1 score is a more reliable measure of algorithm performance. We achieved a mean F1 score, defined as the harmonic mean of the accuracy and repeatability scores, of (77.83±4.90%) using our proposed distance learning approach, compared to (38.72±1.45%) by random guessing.
The Root Mean Square Deviation (RMSE) of the predicted severity level was found to be lower in the MLA approach (28.75±3.81) when compared to random chance assignment (38.93±4.85), suggesting that severity can indeed be assessed to some extent via our metric learning approach.

Discussion

This study indicates that representing IMU measurements as dynamical models (HMMs) and learning optimal pullback Frobenius distances in the space of HMMs was able to classify Parkinsonian gait within a large subset of typically developed adults (TDA), and additionally estimating PD’s severity by means of Nearest neighbour classification.

A study by Barth et al. used Linear Discriminant Analysis (LDA) in order to classify those with Parkinson’s from TDA. The accuracy of this study was based on three activities based upon the MDS-UPDRS Part III (10m walk, heel-toe tapping and foot circling). When combining all three activities, the LDA classifier achieved the best overall accuracy, classifying patients and controls with an accuracy of 88%. Having a much larger number of PwP and TDA samples and a rigorous evaluation protocol, we achieved a higher accuracy in classifying PwP from those who are TDA on our dataset.

Within the current literature only one study uses MLA methods for classification of disease severity10. This study used six different activity classifiers (i.e. k-nearest neighbour, Parzen, Parzen density, Binary decision tree, feed-forward neural network and SVM) to automatically detect the severity of walking-derived bradykinesia according to UPDRS scores. The SVM classifier related the best to the UPDRS scores, with an accuracy range between 70 and 86%.

Our method offers the possibility to use simple gait measure that takes 2 mins to implement in primary care pathways to support general practitioner diagnosis of PD.
Acknowledgements
Competing Interests Statement
Author Contributions
Funding
References (<10)


Figure Legends
Tables
Supplementary