Machine Learning on the Video Basis of Slow Pursuit Eye Movements Can Predict Symptom Development in Parkinson's Patients

Andrzej W. Przybyszewski[1,2(✉)], Stanisław Szlufik[3], Justyna Dutkiewicz[3], Piotr Habela[2], and Dariusz M. Koziorowski[3]

[1] Department Neurology, University of Massachusetts Medical School, Worcester, MA 01655, USA
Andrzej.Przybyszewski@umassmed.edu
[2] Polish-Japanese Institute of Information Technology, 02-008 Warszawa, Poland
{przy,piotr.habela}@pjwstk.edu.pl
[3] Department Neurology, Faculty of Health Science, Medical University of Warsaw, Warsaw, Poland
stanislaw.szlufik@gmail.com, justyna_dutkiewicz@wp.pl, dkoziorowski@esculap.pl

Abstract. We still do not know exactly how brain processes are affected by nerve cell deaths in neurodegenerative diseases such as Parkinson's (PD). Early diagnosis when symptom progressions are precisely monitored may result in improved therapies. In the case of PD, measurements of eye movements (EM) can be diagnostic. In order to better understand their relationship to the underlying disease process, we have performed measurements of slow (POM) eye movements in PD patients. We have compared our measurements and algorithmic diagnoses with doctor's diagnoses. We have used rough set theory and machine learning (ML), to classify how condition attributes predict the neurologist's diagnosis. We have measured pursuit ocular movements (POM) for three different frequencies and estimated patients' performance by gain and accuracy for each frequency. We have tested ten PD patients in four sessions related to combination of medication and DBS treatments. We have obtained a global accuracy in individual patients' UPRDS III predictions of about 80%, based on cross-validation. This demonstrates that POM may be a good biomarker helping to estimate PD symptoms in automatic, objective and doctor-independent way.

Keywords: Neurodegenerative disease · Rough set · Machine learning

1 Introduction

Our approach is to demonstrate an alternative to the mostly used statistical analysis of PD outcomes by using data mining and machine learning (ML) methods. We gave examples that our methods give a more precise description of individual patient's symptoms and development. We may propose an individual treatment adjusted to different

N.T. Nguyen et al. (Eds.): ACIIDS 2015, part II, LNCS 9012, pp. 268–276, 2015.
DOI: 10.1007/978-3-319-15705-4_26

patients that may lead more effectively than now to slowing of symptoms and improvements in quality of life. Our analysis is proposed on the basis of learning algorithms that intelligently process data of each patient in an individual and specific ways.

Our symptom classification method follows the principle of the complex object recognition such as those in visual systems. The ability of natural vision to recognize objects arises in the afferent, ascending pathways that classify properties of objects' parts from simple attributes in lower sensory areas, to more complex ones, in higher analytic areas. The resulting classifications are compared and adjusted by interaction with whole object ("holistic") properties (representing the visual knowledge) at all levels using interaction with descending pathways [1] that was confirmed in animal experiments [2]. These interactions at multiple levels between measurements and prior knowledge can help to differentiate individual patient's symptoms and response treatments variability in a way similar to a new, complex object inspection [3, 4]. Machine learning algorithms for analyzing subtle signal variations will hopefully lead to better analysis of individual patients' conditions. As it was demonstrated in [1, 3, 4] properties of the primates visual system can be well described by rough set theory, therefore we have applied the same concept to knowledge discovery from symptoms in PD.

Based on their experience, intuition and at least partly subjective measurements neurologists are giving diagnosis of individual patients. They use "Golden Standard" by estimation values of the Hoehn and Yahr scale and the UPDRS (Unified Parkinson's Disease Rating Scale). As different doctors are not always in the precise way perform exactly same procedure their diagnosis are partially subjective that may lead to different treatments. We propose to formalize the whole process and use the neurologist's diagnosis as decision attributes and their and our doctor-independent measurements as condition attributes.

2 Methods

Our experiments were performed on ten Parkinson Disease (PD) patients who had undergone the Deep Brain Stimulation (DBS) surgery mainly for treatment of their motor symptoms. They were qualified for the surgery and observed postoperatively in the Dept. of Neurology and got surgical DBS implementation in the Institute of Neurology and Psychiatry [5]. We conducted horizontal POM (pursuit ocular movement - as explained below) measurements in ten PD patients during four sessions designated as S1: MedOffDBSOff, S2: MedOffDBSOn, S3: MedOnDBSOff, S4: MedOnDBSOn. During the first session (S1) the patient was off medications (L-Dopa) and DBS stimulators was OFF; in the second session (S2) the patient was off medication, but the stimulator was ON; in the third session (S3) the patient was after his/her doses of L-Dopa and the stimulator was OFF, and in the fourth session (S4) the patient was on medication with the stimulator ON. Changes in motor performance, behavioral dysfunction, cognitive impairment and functional disability were evaluated in each session according to the UPDRS. The pursuit (POM) was recorded by head-mounted saccadometer (Ober Consulting, Poland). We have used an infrared eye track system coupled with a

head tracking system (JAZZ-pursuit – Ober Consulting, Poland) in order to obtain high accuracy and precision in eye tracking and to compensate possible subjects' head movements relative to the monitor. Thus subjects did not need to be positioned in an unnatural chinrest.

A patient was sited at the distance of 60-70 cm from the monitor with head supported by a headrest in order to minimize head motion. We measured slow eye movements in response to a light spot with horizontal sinusoidal movements (three frequencies: 0.125, 0.25, 0.5Hz) from 10 deg to the left to 10 deg to the right (the exact rage of the spot amplitude (in degrees) depends on the patient's distance from the screen). At first the patient has to fixate eyes on the spot in the middle marker (0 deg) the spot was placed in 10 deg to the left and 10 deg to the right for the calibration. In the next step, patients had to look at the targets (small square) and follow its sinusoidal, horizontal movement.

In each test the subject had to perform 4 periods of POM with low and 10 with higher frequencies in Med-off (medication off) within two situations: with DBS off (S1) and DBS on (S2). In the next step the patient took medication and had a break for one half to one hour, and then the same experiments were performed, with DBS off (S3) and DBS on (S4). In this work we have analyzed POM data using the following population parameters averaged for both eyes: gain (eye movement amplitude/sinus amplitude) and accuracy (difference between sinusoid and eye positions) for three different frequencies.

2.1 Theoretical Basis

We represent our data in the form of information system that is also called the decision table. We define such an information system (after Pawlak [6]) as a pair $S = (U, A)$, where U, A are nonempty finite sets called the *universe of objects* and the *set of attributes*, respectively. If $a \in A$ *and* $u \in U$, the value $a(u)$ is a unique element of V (where V is a value set).

The *indiscernibility relation* of any subset B of A or $IND(B)$, is defined [6] as follows: $(x, y) \in IND(B)$ or $xI(B)y$ if and only if $a(x) = a(y)$ for every $a \in B$, where $a(x) \in V$. $IND(B)$ is an equivalence relation, and $[u]_B$ is the equivalence class of u, or a *B-elementary granule*. The family of all equivalence classes of $IND(B)$ will be denoted $U/I(B)$ or U/B. The block of the partition U/B containing u will be denoted by $B(u)$.

We define a *lower approximation* of symptoms set $X \subseteq U$ in relation to a symptom attribute B as $\underline{B}X = \{u \in U: [u]_B \subseteq X\}$, and the *upper approximation* of X as $\overline{B}X = \{u \in U: [u]_B \cap X \neq \phi\}$. It means that, symptoms are classified into two categories (sets). The lower approximation set X has the property that all symptoms with certain attributes are part of X, and the upper movement approximation set has property that only some symptoms with attributes in B are part of X (see [5]). The difference of $\overline{B}X$ *and* $\underline{B}X$ is defined as the boundary region of X i.e., $BN_B(X)$. If $BN_B(X)$ is empty set than X is *exact* (*crisp*) with respect to B; otherwise if $BN_B(X) \neq \phi$ and X is not *exact* (i.e., it is *rough*) with respect to B. We say that the B-lower approximation of a given set

X is union of all *B-granules* that are included in *X*, and the *B*-upper approximation of *X* is of the union of all *B-granules* that have nonempty intersection with *X*.

The system *S* will be called a decision table $S = (U, C, D)$ where *C* is the condition and *D* is the decision attribute [6]. In the table below (Table 2), as an example, the decision attribute *D,* based on the expert opinion*,* is placed in the last column, and condition attributes measured by the neurologist, are placed in other columns. One can interpret each row in the table as a rule. As the number of rules is same as the number of rows, and each row is related to different measurements, these rules can have many particular conditions. We would like to describe different symptoms in different patients by using such rules. On the basis of such rules, using the modus ponens principle we wish to find universal rules to relate symptoms and treatments in different patients [6]. As symptoms even for the same treatments are not always the same; our rules must have certain "flexibility", or granularity, which can be interpreted as the probability of finding certain symptoms in a group of patients under consideration. The granular computation simulates the way in which neurologists interact with patients. This way of thinking relies on the ability to perceive a patient's symptoms under various levels of granularity (i.e., abstraction) in order to abstract and consider only those symptoms that serve to determine a specific treatment and thus to switch among different granularities. By focusing on different levels of granularity, one can obtain different levels of knowledge, as well as a greater understanding of the inherent knowledge structure. As one of us has demonstrated [1, 2] that the visual system is using the granular computing in object recognition, we suggest that this approach is essential for human intelligent.

We define the **reduct** $B \subset A$. The set B is a reduct of the information system if $IND(B) = IND(A)$ and no proper subset of *B* has this property. In case of decision tables decision reduct is a set $B \subset A$ of attributes which cannot be further reduced and $IND(B) \subset IND(d)$. A decision rule is a formula of the form $(a_{i1} = v_1) \wedge ... \wedge (a_{ik} = v_k) \Rightarrow d = v_d$, where $1 \leq i_1 < ... < i_k \leq m$, $v_i \in Va_i$. Atomic subformulas $(a_{i1} = v_1)$ are called conditions. In this way, we can replace the original attribute a_i with new, binary attributes, which indicate whether actual attribute value for an object is greater or lower than *c* (see [7]), we define *c* as a cut. Thus a cut for an attribute $a_i \in A$, with V_{ai} will be a value $c \in V_{ai}$. A template of *A* is a propositional formula $v_i \in V_{ai}$. A generalized template is a formula of the form $\wedge(a_i \in T_i)$ where $T_i \subset V_{ai}$. An object satisfies (matches) a template if for every attribute a_i we have $a_i = v_i$ where $a_i \in$ A. The template is a method to split the original information system into two distinct sub-tables. One of these sub-tables consists of the objects that satisfy the template, while the second contains all others. A *decomposition tree* is defined as a binary tree, whose every internal node is labeled by some template and external node (leaf) is associated with a set of objects matching all templates in a path from the root to a given leaf [8].

In a second test we have divided our data into two or more subsets. By training on all but one of these subsets (the training set) using machine learning (ML), we obtained classifiers that when applied to the remaining (test) set gave new numerical decision attributes, well correlated with neurologist decision attributes (based on a confusion matrix).

3 Results

The patients' mean age was 58.3±9.3(SD) years, mean disease duration was 10.9±1.6 years, mean UPDRS (related to all symptoms): S1: 59.4±16.2 S2: 29.9±13.3; S3: 51.2±14.4; S4: 18.2±11.4; mean UPDRS III (related only to motor symptoms): S1: 43.5±12.7 S2: 20.4±7.9; S3: 35.3±11.1; S4: 9.6±5.9.

The differences between UPDRS/UPDRS III: S1-S2, and S1-S4 were statistically significant ($p< 0.001$) and S1-S3 was not statistically significant. Our slow eye movement (POM) measurements did change significantly with the session number: gain - for slow/medium/fast sinusoids were: S1: 1.06±0.1/0.96±0.2/0.83±0.2 S2: 1.03±0.1/0.97±0.2/0.86±0.1 S3: 1.05±0.1/0.99±0.1/0.94±0.1 S4: 1.00±0.1/0.96±0.1/0.87±0.1. Accuracy – as sum of normalized differences between stimulus and eye position - for slow/medium/fast sinusoids were: S1: 0.70±0.13/0.62±0.16/0.54±0.18 S2: 0.67±0.18/0.68±0.16/0.61±0.20 S3: 0.73±0.13/0.70±0.18/0.63±0.18 S4: 0.78±0.13/0.76±0.15/0.66±0.18.

3.1 Rough Set and Machine Learning Approach

As described above we have used the RSES 2.2 (Rough System Exploration Program) [8] in order to find regularities in our data. At first our data was placed in the information table as originally proposed by Pawlak [6].

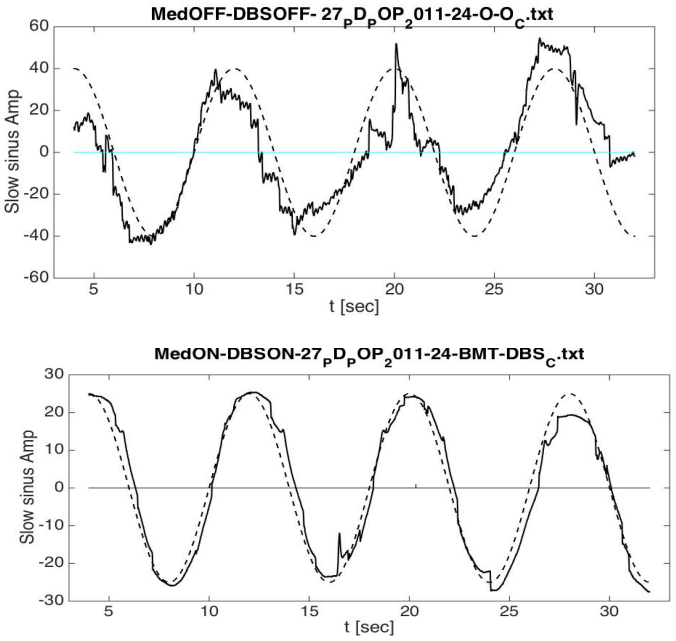

Fig. 1. Experimental recordings of POM from patient #27 before (upper part) and after (lower plot) medications and DBS treatments

Table 1. Extract from the information table

P#	age	sex	t_d	S#	UPDRS	HYsc	gxss	gxms	gxfs	accss	accms	accfs
28	54	1	8	1	58	2.0	0.94	1.04	0.97	0.71	0.86	0.81
28	54	1	8	2	40	1.0	1.04	0.98	0.93	0.91	0.93	0.86
28	54	1	8	2	40	1.0	1.17	1.07	0.91	0.82	0.90	0.67
28	54	1	8	4	16	1.0	1.08	1.00	0.90	0.86	0.89	0.69
38	56	0	11	1	49	2.5	0.90	0.94	0.94	0.73	0.76	0.66
38	56	0	11	2	22	1.5	1.04	1.03	0.93	0.89	0.85	0.76
38	56	0	11	3	37	2.5	0.99	1.01	1.03	0.83	0.81	0.69
38	56	0	11	4	12	1.5	1.08	1.11	1.03	0.81	0.77	0.76

The full table has 14 attributes and 40 objects (measurements). In the Table 1 are values of 11 attributes for two patient: P# - patient number, age – patient's age, sex – patient's sex: 0 - female, 1 – male, t_d – duration of the disease, S# - Session number, UPDRS – total UPDRS, HYsc – Hoehn and Yahr scale all measured by the neurologist and POM measurements: gxss - gain for slow sinus; gxms - gain for slow sinus; gxfs - gain for slow sinus; accss - accuracy for slow sinus; gxms - accuracy for medium sinus; gxfs - accuracy for fast sinus;

In the next step, we have performed reduction of attributes (see reduct in the Method section) to a minimum number of attributes describing our results. We have also created a discretization table: where single values of measurements were replaced by their range (as describe in the Method section on cut sets). As the result we have obtained the decision table (Table 2 –see below).

Table 2. Part of the decision discretized-table

Pat#	age	accfs	Ses#	HYsc	SchEng	gxms	gxfs	UPDRS III
'28	*	"(0.75,Inf)"	1	*	"(-Inf, 85)"	"(1.04,Inf)"	"(0.845,Inf)"	"(28.0,Inf)"
'28	*	"(0.75,Inf)"	2	*	"(-Inf, 85)"	"(0.97,1.04)"	"(0.845,Inf)"	"(16.5,28.0)"
28	*	"(0.39,0.75)"	3	*	"(-Inf, 85)"	"(1.04,Inf)"	"(0.845,Inf)"	"(16.5,28.0)"
:28	*	"(0.39,0.75)"	4	*	"(85, Inf)"	"(0.97,1.04)"	"(0.845,Inf)"	"(-Inf,16.5)"
'38	*	"(0.39,0.75)"	1	*	"(-Inf, 85)"	"(-Inf,0.97)"	"(0.845,Inf)"	"(28.0,Inf)"
'38	*	"(0.75,Inf)"	2	*	"(-Inf, 85)"	"(0.97,1.04)"	"(0.845,Inf)"	"(-Inf,16.5)"
'38	*	"(0.39,0.75)"	3	*	"(-Inf, 85)"	"(0.97,1.04)"	"(0.845,Inf)"	"(-Inf,16.5)"
'38	*	"(0.39,0.75)"	4	*	"(85, Inf)"	"(1.04,Inf)"	"(0.845,Inf)"	"(-Inf,16.5)"

In the first column is the patient's number, in the second the patient's age not important (*); next was accfs – accuracy for fast sinus freq; Ses# -Session number,

Hoehn and Yahr scale were not considered important (stars); SchEng -Schwabe England scale; gxms – gain got medium sinus; gxfs – gain for fast freq. sinus and UPDRS III that was divided into different ranges: above 28, 16.5 to 28,, and below 16.5 (the last column). On the basis of this decision table we can write the following rule:

('Pat'=28)&('accfs'="(0.75,Inf)")&('Sess'=1)&('SchEng'="(-Inf,85)")&('gxms' ="(1.04,Inf)")& ('gxfs' ="(0.845,Inf)") => ('UPDRS III'="(28.0,Inf)") (1)

We read this formula above (eq. 1), as stating that each row of the table (Table 1) can be written in form of this equation (eq. 1). It states that if we evaluate patient #28 *and* with accfs above 0.75 *and* in session #1 *and* with Schwabe England scale below 85. *and* gxms (gain fom medium freq. sinus) above 1.04 *and* … *and* gxfs above 0.845 *then* patient's UPDRS is above 28.

These equations are parts of a data mining system bases on rough set theory [6]. We have tested our rule using the machine-learning concept. Randomly dividing our data into 4 groups, we took 3 groups as training set and tested the fourth. By changing groups belonging to the training and test sets, we have removed the effect of accidental group divisions. The results of each test were averaged – thus we have performed a 4-fold cross-validation. The results are gives as a confusion matrix (Table 3). As a machine-learning algorithm we have used the decomposition tree (see Methods).

Table 3. Confusion matrix for different session numbers (S1-S4)

		Predicted			
		28.0, Inf	**16.5, 28.0**	**-Inf, 16.5**	**ACC**
Actual	**28.0, Inf**	0.5	0.5	0.0	**0.33**
	16.5, 28.0	0.25	0.0	0.25	**0.0**
	-Inf, 16.5	0.0	0.25	2.25	**0.67**
	TPR	**0.5**	**0.0**	**0.67**	

TPR: True positive rates for decision classes, ACC: Accuracy for decision classes: the global coverage was 0.4, the **global accuracy was 0.774**, coverage for decision classes: 0.25, 0.3, 0.6.

Another question that result is, whether EM can help to estimate possible effects of different treatments in individual patients? In order to demonstrate an answer, we have removed EM measurements and added other typically measured attributes such as: the Schwab and England ADL Scale, and UPDRS III and UPDRS IV to the decision table and tried to predict the effects of different treatments as represented by sessions 1 to 4 (medication and stimulation effects- results are in Table 4) and compared them with predictions based on POM (results are in Table 5).

We have performed the same procedures once more to test results of patients' eye movement influence on our predictions.

Table 4. Confusion matrix for different session numbers (S1-S4)

		Predicted				
		1	**2**	**3**	**4**	**ACC**
Actual	**1**	0.5	0.0	0.5	0.0	**0.3**
	2	0.0	0.5	0.0	0.3	**0.4**
	3	0.8	0.0	0.2	0.0	**0.2**
	4	0.0	0.5	0.0	0.5	**0.4**
	TPR	**0.3**	**0.3**	**0.2**	**0.4**	

TPR: True positive rates for decision classes, ACC: Accuracy for decision classes: the global coverage was 0.64, the **global accuracy was 0.53**, coverage for decision classes: 0.5, 0.5, 0.75, 0.7.

Table 5. Confusion matrix for different session numbers (S1-S4)

		Predicted				
		1	**2**	**3**	**4**	**ACC**
Actual	**1**	1.75	0.0	0.0	0.0	**0.75**
	2	0.0	0.25	0.0	0.5	**0.25**
	3	0.0	0.0	0.5	0.0	**0.5**
	4	0.0	0.25	0.0	0.75	**0.33**
	TPR	**0.7**	**0.7**	**0.6**	**0.25**	

TPR: True positive rates for decision classes, ACC: Accuracy for decision classes, the global coverage was 0.45; the **global accuracy was 0.795**; coverage for decision classes: 0.58, 0.21, 0.38, 0.42.

In summary, two last results have demonstrated that adding eye movement (EM) results to classical measurements performed by the most neurologists, can result in improved predictions of disease progression measured, as measured by improvement in global accuracy from 0.5 to 0.8. The EM measurements may also partly replaces neurological measurements such as the UPDRS, as global accuracy of the total UPDRS predictions taken from EM data was 0.77 for the above 10 PD patients.

4 Discussion

In current therapeutic protocols, even with the large numbers of approaches and clinical trials, there have still been few conclusive results on therapeutic identification and measurement of PD symptoms. We have given an example comparing classical neurological diagnostic protocols with a new approach. The main difference between

these types of measures is in their precision and objectivity. Our approach is doctor-independent and can be performed automatically. In the near future it may help in transforming some hospital-based to home-based treatments. In this scenario it will be possible to measure patient symptoms at home, and send these for consultation by neurologists.

5 Conclusions

We have presented a comparison of classical statistical averaging methods for PD diagnosis with rough set (RS) approaches. We used processed neurological data from PD patients in four different treatments and we have plotted averaged effects of the medication and brain stimulation in individual patients. As these effects are strongly patient dependent they could not give enough information to predict new patient's behavior. The RS and ML approaches are more universal giving general rules for predicting individual patient responses to treatments as demonstrated in UPDRS predictions.

Acknowledgements. This work was partly supported by projects Dec-2011/03/B/ST6/03816 and NN 518289240 from the Polish National Science Centre.

References

1. Przybyszewski, A.W.: The neurophysiological bases of cognitive computation using rough set theory. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 287–317. Springer, Heidelberg (2008)
2. Przybyszewski, A.W., Gaska, J.G., Foote, W., Pollen, D.A.: Striate cortex increases contrast gain of macaque LGN neurons. Visual Neuroscience **17**, 1–10 (2000)
3. Przybyszewski, A.W.: Logic in Visual Brain: Compute to Recognize Similarities: Formalized Anatomical and Neurophysiological Bases of Cognition. Review of Psychology Frontier **1**, 20–32 (2010) (open access)
4. Przybyszewski, A.W.: Logical rules of visual brain: From anatomy through neurophysiology to cognition. Cognitive Systems Research **11**, 53–66 (2012)
5. Pizzolato, T. Mandat, T.: Deep Brain Stimulation for Movement Disorders. Frontiers in Integrative Neuroscience **6**, 2 (2012) doi:10.3389/fnint.2012.00002 (Published online January 25, 2012)
6. Pawlak, Z.: Rough sets: Theoretical aspects of reasoning about data. Kluwer, Dordrecht (1991)
7. Bazan, J., Nguyen, H.S., Nguyen, T.T., Skowron A., Stepaniuk, J.: Desion rules synthesis for object classification. In: Orłowska, E. (ed.) Incomplete Information: Rough Set Analysis, pp. 23–57. Physica – Verlag, Heidelberg (1998)
8. Bazan, J., Szczuka, M.S.: RSES and RSESlib - a collection of tools for rough set computations. In: Ziarko, W.P., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 106–113. Springer, Heidelberg (2001)