Fuzzy Rough Sets Theory Applied to Parameters of Eye Movements Can Help to Predict Effects of Different Treatments in Parkinson's Patients

Anna Kubis[1]([✉]), Artur Szymański[2],
and Andrzej W. Przybyszewski[2,3]

[1] Faculty of Applied Mathematics, AGH University of Science and Technology,
A. Mickiewicza 30, 30-059 Kraków, Poland
anna.kubis@fis.agh.edu.pl
[2] Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland
artur.szymanski@pja.edu.pl
[3] Department of Neurology, University of Massachusetts Medical School,
65 Lake Av., Worcester, MA 01655, USA
andrzej.przybyszewski@umassmed.edu

Abstract. Parkinson (PD) is the second most common neurodegenerative disease (ND) with characteristic movement disorders. There are well defined standard procedures to measure disease stage (Hohen Yahr scale), progression and effects of treatments (UPDRS – unified Parkinson Disease Rate Scale). But these procedures can only be performed by experienced neurologist and they are partly subjective. The purpose of our work was to test objective and non-invasive method that may help to estimate disease stage by measuring fast and slow eye movements (EM). It was demonstrated earlier that EM changes in PD. We have measured reflexive saccades (RS) and slow pursuit ocular movements (POM) in four sessions related to different treatments. With help of fuzzy rough sets theory (FRST) we have related measurements with expert's opinion by generalizing experimental finding by *fuzzy rules*. In order to test our approach, we have divided our measurements into training and testing sets. In the second test, we have removed expert's decisions and predicted them from the training set in two situations: on the basis of only classical neurological measurements and on the basis of EM measurements. We have observed, on 12 PD patients basis, an increase in predictions accuracy when eye movements were included as condition attributes. Our results with help of the FRST suggest that EM measurements may become an important diagnostic tool in PD.

Keywords: Deep brain stimulation (DBS) · Reducts · Information table · Decision rules · Rough sets

1 Introduction

The brain's deeper computational properties are still not well understood. We are even not sure if brain computations are more powerful than the Turing machine and such models as ARNN (analog recurrent neural networks) or coupled nonlinear oscillators are appropriate [1, 2]. For example, we do not know exactly how brain processes are affected by nerve cell deaths in the neurodegenerative diseases (ND) such as Parkinson or Alzheimer. It is well documented, however, that the disease starts long before the observed first symptoms and individual pathological mechanisms have a large spectrum. In Parkinson's, for example, the first motor symptoms are observed when 70 −80 % cells in responsible structure (substantia nigra) are dead and once cells are dead there is no chance for their recovery.

We can register symptoms of ND such as motor and/or mental disorders (dementias) and even provide symptomatic relief, though the structural effects of these are in most cases not yet understood. Fortunately, with early diagnosis there are often many years of disease progression with symptoms that, when they are precisely monitored, may result in improved therapies.

One of the purposes of this work is to try to extract knowledge from symptoms in order to model possible mechanisms of disease progression and adjust therapies in timely precise matter.

The majority of neurologists use the standard statistical methods to analyze the results of PD patients' treatment. As every patient suffers from PD in a different way and reacts differently to the treatment, averaging methods can lead to the confusing results. Therefore, in continuation of [3, 4], we propose to extend statistical analysis to data mining techniques in order to adjust PD treatment to an individual patient. Our method is based on fuzzy rough sets theory application as this approach should better fit to predict partly noisy and continuous medical measurements than previously proposed rough sets theory [3, 4].

As PD progression biomarker we have used measurements of eye movements. It is well established on the animal experimental basis that the basal ganglia are involved in the eye movement's control (see review [5]). It was also demonstrated on human subjects that fast (saccades) and also slow (pursuit ocular movements) eye movements are affected in Parkinson's diseases [6, 7].

Generally, different treatments are based on the UPDRS (Unified Parkinson's Disease Rating Scale) measurements, in particular on UPDRS II (activity of daily living), UPDRS III (examination of motor symptoms), UPDRS V (modifies Hoehn and Yahr staging – stage of the disease) and UPDRS VI (Schwab and England activities of daily living scale). As these measurements are strongly doctor dependent and partly subjective, we propose to use the eye movement (EM) as an individual doctor independent measure to improve diagnosis and objectivity. In the consequence, in our analysis in addition to standard neurological measurements, we have added EM parameters as condition attributes, doctors' expertise as the decision attribute and placed them in the decision table [8]. As the data in the table were related to different treatments, our purpose was to use the data mining techniques to estimate and to predict effectiveness of different therapies related to individual patients.

2   Methods

We have performed our analysis on PD data used earlier in [3, 4]. All of the 12 patients had implanted electrodes in the subthalamic nucleus that is a standard procedure in advanced Parkinson's. As number of PD patients is relatively small, results of this study are preliminary. The measurements were conducted in four sessions (S1−S4): in the first session (S1) patients were off medications (L-Dopa) and DBS stimulators was OFF; in the second session (S2) patient were off medication, but the stimulator was ON; in the third session (S3) patients were after his/her doses of L-Dopa and stimulator was OFF, and in the fourth session (S4) patients were on medication with the stimulator ON. The data set consisted of the estimation of the disease advancement made during the medical interview (expressed by Unified Parkinson Disease Rate Scale - UPDRS) related to changes in motor performance, behavioral dysfunction, cognitive impairment and functional disability, and EM measurements. We have evaluates saccadic and slow pursuit eye movements. The EM were 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 the EM measurements patient was sitting at the distance of 60−70 cm from the monitor with head supported by a headrest in order to minimize head motion. We measured fast eye movements in response to a light spot switched on and off, which moved horizontally from the straight eye fixation position (0 °) to 10 ° to the left or 10 ° to the right after arbitrary time ranging between 0.5–1.5 s. When the patient fixated eyes on the spot in the middle marker (0 °) the spot then changed color from white to green, indicating a signal for performance of RS (reflexive saccades); or from white to red meaning a signal for performing AS (antisaccades) – not evaluated in this study. Then the central spot was switched off and one of the two peripheral targets, selected at random with equal probability, was illuminated instead (non-overlapping test). Patients had to look at the targets and follow them as they moved in the RS task. After making a saccade to the peripheral target, the target remained on for 0.1 s after which another trial was initiated. In each test the subject had to perform 10 RS and 10 AS in a row 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). Slow EM – pursuit ocular movements (POM) were measured in response to a light spot with horizontal sinusoidal movements (with slow (0.125 HZ), medium (0.25 Hz) and fast (0.5 Hz) frequencies), placed from 10 ° to the left to 10 ° to the right. POM measurements were performed in four different sessions in similar procedures as described above for RS measurements.

In this work we have analyzed only RS data using the following parameters: averaged for both eyes: delay (RS latency), amplitude (RS amplitude), duration (RS duration), velocity (RS velocity). We have analyzed POM data using the following parameters averaged for both eyes: gain (eye movement amplitude/sinus amplitude) and accuracy (difference between sinusoid and eye positions) for three different frequencies. More details can be found in [3, 4].

3 Theoretical Basis

Our data were represented as a decision table. In the rows we put the measurements' values for respective patients during each single session. As columns we use patient's number, patient's age, session number, estimations of UPDRS, Schwab and England and Hoehn and Yahr scales and EM measurements: RS parameters and slow, medium and fast sinus parameters for POM.

As fuzzy rough set theory (FRST) is an extension of rough set theory (RST) [8] we define here a similarity or tolerance relation [9, 10] instead of crisp equivalence. The tolerance relation $R_a(x, y)$ determines the discernibility between the values of the specific attribute for a pair of observation. There are several means to describe this relation $R_a(x, y)$ as presented below after [8–12]:

$$R_a(x, y) = 1 - \frac{|a(y) - a(x)|}{|a_{min} - a_{max}|} \tag{1}$$

In this way, the value of tolerance relations is directly proportional to the absolute value of the difference between the attribute's values for the two observations.

$$R_a(x, y) = e^{\frac{-(a(y)-a(x))^2}{2\sigma_a}} \tag{2}$$

where σ_a stands for standard deviation for the given attribute a. This equation includes standard deviation of the data, therefore in most cases it is more sensible for the behavior of the data than Eq. (1) mentioned above.

$$R_a(x, y) = e^{\frac{-\|a(y)-a(x)\|^2}{d}} \tag{3}$$

where d is a positive number. In our case, we take absolute value as a norm and variance σ_a in place of d.

In the next step, we have normalized the differences between each pair of conditional attributes' values. For this purpose, we have used a t-norm, marked τ. For a given pair of attributes a and b we get $R_{\{a,b\}}(x, y) = \tau(R_a(x, y), R_b(x, y))$. In order to get the value of the relations for the whole set of conditional attributes B, it is enough to normalize the difference between the first pair and the successive element and then by recurrence the difference between the value for the set got at the preceding step and a successive added element: $R_{\{a,b,c\}}(x, y) = \tau(R_{\{a,b\}}(x, y), R_c(x, y))$. The two most commonly used t-norms are: t.cos and Łukasiewicz t-norm, described respectively by equations:

$$R_{\{a,b\}}(x, y) = \max\{0, R_a(x, y) \cdot R_b(x, y) - \sqrt{1 - R_a(x, y)} \cdot \sqrt{1 - R_b(x, y)}\}, \tag{4}$$

$$R_{\{a,b\}}(x, y) = \max\{0, R_a(x, y) + R_b(x, y) - 1\}. \tag{5}$$

Tolerance relation defined by (1) is transitive in both t-norms, while tolerance relations defined with (2, 3) are transitive only with t-norm [9].

In case of modelling the difference between values for the decision attribute, we usually use the relation of identity: $R_d(x, y) = \begin{cases} 0, x = y \\ 1, x \neq y \end{cases}$.

In FRS concept, for the sets *(U, B)* of observations and condition attributes we define *B*-lower and *B*-upper approximations separately for every observation *x*. For each of the observations *x* we define *B*-lower approximation as: $(R_B \downarrow X)(x) = \inf_{y \in U} I(R_B(x, y), X(y))$, where *I* is an implicator [9]. The *B*-lower approximation for the observation *x* is then the set of observations which are the most similar to observation *x* and it can predict the decision attribute with the highest confidence, based on conditional attributes *B*.

The *B*-upper approximation is defined by $(R_B \uparrow X)(x) = \sup_{y \in U} \tau(R_B(x, y), X(y))$. Then, in fuzzy rough sets approach, the *B*-upper approximation is a set of observations for which the prediction of decision attribute has the smallest confidence.

Another term used in further explanations is positive region for an element y. The fuzzy B-positive region is a fuzzy set in the set *U* that contains each observation x to the extent that all objects with approximately equal values for the set of conditional attributes *B* have equal values for decision attribute. Formally after [9]: $POS_B(y) = \bigcup (R_B \downarrow R_d x)(y)$.

The predictive ability for d of the set of conditional attributes *B* is reflected in the degree of dependency defined as $\gamma_B = \frac{|POS_B|}{|POS_{A\backslash\{d\}}|} = \frac{\sum_{x \in U} POS_B(x)}{\sum_{x \in U} POS_{A\backslash\{d\}}(x)}$. If there does not exist any other subset B′ of B such that $POS_{B'} = POS_{A\backslash\{d\}}$, B is called a decision reduct.

The rules in FRST approach are constructed from tolerance classes and corresponding decision concepts. A ready fuzzy rule will be a triple *(B, C, D)*, where *B* is a set of conditional attributes that appear in the rule, *C* stands for fuzzy tolerance class of object and *D* stands for decision class of object.

Apparently, many terms are defined differently in rough sets (applied and described in [3, 4]) and fuzzy rough sets approach, e.g. in RST upper approximation is a global term, defined for the whole data set while in FRST we define upper approximation separately for each element. Those sets are also larger in fuzzy method, as they contain observations that are not necessarily identical with the observation for which we define the upper approximation. As a consequence of this difference, in most cases we get close to 100 % of class coverage for predictors in FRST approach, while the coverage in RST is usually much lower.

4 Results

Below are examples of decision tables that include fast eye movements - reflexive saccades: RS (Table 1) and slow, pursuit eye movements - POM (Table 2) parameters.

Table 1. A part of the decision table for the first experiment including RS

Pat	Age	Sess	UPDRS III	Schwab EngSc	Scc Durat	SccLat	SccAmp	SccVel	UPDRS Total
11	58	1	45	60	52	259	15	552,3	75
11	58	2	17	90	52	204	14	549,8	32
11	58	3	32	60	46	315	12	514	62
11	58	4	10	90	54	239	11	371,7	25

Pat - patient's number, age - patient's age, Sess - session number, UPDRS III - motor tests, SchwabEngSc –Schwab & England activity, SccDurat - RS duration, SccLat - RS latency, SccAmp - RS amplitude, SccVel - RS velocity, UPDRS Total - sum of a UPDRS I to VI.

On the basis of Table 1 we understand rules as (the last column is the decision attribute) for the first row:

$$('Pat' = 11)\&('age' = 58)\&('Sess' = 1)\&\ldots = > ('UPDRS\,III' = 45) \quad (6)$$

The rule should be read as follows: if for patient #11 and his/her age 58 and session S1 and value equal to 45 and … then his/her UPDRS III value is 45. We get such rules separately for each of the rows of the decision table. The main purpose of our analysis is to reduce number and increase universality of these rules.

In order to create fuzzy rules, we have used the algorithm called Hybrid Fuzzy-Rough Rule Induction and Feature Selection and described in detail in [13–15]. In the mentioned algorithm feature selection (a process of finding a subset of attributes which represent the same information as the complete feature set) and rule induction are performed simultaneously.

Table 2. A part of the decision table including POM

Pat	Age	Sess	HY scale	Schwab EngSc	gxss	gxms	gxfs	accss	accms	accfs
11	58	1	4	60	1.0583	0.9327	0.8487	0.8863	0.8151	0.7434
11	58	2	1	90	1.0077	0.9145	0.8823	0.5768	0.658	0.6929
11	58	3	4	60	1.0336	1.0046	0.9498	0.6864	0.652	0.4858
11	58	4	1	90	1.0588	1.0262	0.9408	0.7838	0.8019	0.8384

Pat- patient's number, age - patient's age, Sess - session number, HYscale - Hoehn and Yahr's scale, SchwabEngSc- as above, gxss/gxms/gxfs - gain for slow/medium/fast sinus, accss/accms/accfs - accuracy for slow/medium/fast sinus.

The rules determining UPDRS III are important in prediction of PD symptoms while the rules for session numbers are crucial in measuring the effects of different treatments. In order to predict results from new patients, we have performed the *test-and-train scenario* (e.g. [8]). For this purpose we divide the data set into two parts: training set, containing 75 % of the data and testing set, composed of the remaining 25 % that we have tested. We have removed decision attributes from the test set and compared them with attributes values obtained from our rules.

As the test-and-train scenario strongly depends on which part of our measurements was taken as training and which part was tested. In order to make the result possibly generalized, we have divided our experimental set into 4 subsets (4-fold-test). Then we treated each of them separately as a testing set, using the sum of other sets as a training set. The mean of four predictions gave the final measure of the accuracy.

In order to measure the effects of the treatment, we performed the prediction of the session numbers as a decision attribute. In the first step, as other attributes we used patient's number, age, results of UPDRS III, UPDRS IV and UPDRS Total, result in Hoehn and Yahr's scale and in Schwab and England's scale.

To make the prediction, we have used the RoughSets package in R environment. We checked the results of the prediction using different tolerance and t-norm definitions (Table 3).

Table 3. Global accuracy for different parameters chosen for the prediction of session numbers (S1−S4) without EM attributes

tolerance relation	t-norm	global accuracy
eq.(1)	eq.(5)	0.33
eq.(1)	eq.(4)	0.38
eq.(2)	eq.(4)	0.29
eq.(3)	eq.(4)	0.29

We chose then Eqs. 1 and 4 as parameters for our prediction. Its results are presented in the confusion matrix below (Table 4).

Table 4. Confusion matrix for different session numbers (S1−S4) without EM attributes

Predicted

		1	2	3	4	ACC
	1	0.5	0.5	0.25	0	**0.4**
	2	0	2	0.25	0	**0.89**
Actual	**3**	0.75	0.75	0	0	**0**
	4	0.25	0.75	0	0	**0**
	TPR	**0.33**	**0.5**	**0**	**NaN**	

TPR: True positive rates for decision classes, ACC: Accuracy for decision classes. Class coverage for predictors: 1 and global coverage = 1, and **global accuracy = 0.42**.

In order to compare how eyes movements parameters change the diagnostic ability of the data set, in the second step we have used the parameters of eyes movements: POM gain and accuracy for medium sinus (Table 5).

Table 5. Global accuracy for prediction of session numbers (S1−S4) including POM attributes

tolerance relation	t-norm	global accuracy
eq.(1)	eq.(5)	0.46
eq.(1)	eq.(4)	0.55
eq.(2)	eq.(4)	0.46
eq.(3)	eq.(4)	0.5

We chose then Eqs. (1) and (4) as parameters for our prediction. Its results are presented in the confusion matrix below (Table 6).

Table 6. Confusion matrix for different session numbers (S1−S4) including POM

Predicted

		1	2	3	4	ACC
	1	1	0.5	0	0	**0.67**
	2	0.25	0.75	0	0.25	**0.6**
Actual	3	0	0.25	0.75	0.5	**0.5**
	4	0.5	0	0.5	0.75	**0.42**
	TPR	**0.57**	**0.5**	**0.6**	**0.5**	

TPR: True positive rates for decision classes, ACC: Accuracy for decision classes. Class coverage for predictors: 1 and global coverage = 1, and **global accuracy = 0.55**.

In order to predict individual patient's symptoms related to different treatments, we made prediction of the UPDRS III. To estimate the global accuracy for the predictions of UPDRS attributes, we decided to recognize the prediction as accurate if it does not differ from the actual values from more than 20 % of values range.

As our purpose was to find if RS (saccade) attributes are significant, we began with prediction of UPDRS III using classical neurological measures but without UPDRS total and without EM parameters. Below, in Table 7 we gave results of global accuracies using different parameters of tolerance and t-norm. The best result gave Eqs. 1 and 6: the global accuracy was 46 %.

In the next step, we have tested results of UPDRS III prediction using in addition to standard neurological data (without UPDRS total) also RS duration and amplitude. The best result - global accuracy of 63 % - was obtained for Eqs. 2 and 5 (Table 8).

Table 7. Global accuracy for different parameters chosen for the prediction of UPDRS values, not including any eye movements parameters

tolerance relation	t-norm	global accuracy
eq.(1)	eq.(5)	0.46
eq.(1)	eq.(4)	0.33
eq.(2)	eq.(4)	0.38
eq.(3)	eq.(4)	0.38

Table 8. Global accuracy for the UPDRS III prediction including RS

tolerance relation	t-norm	global accuracy
eq.(1)	eq.(5)	0.58
eq.(1)	eq.(4)	0.54
eq.(2)	eq.(4)	0.63
eq.(3)	eq.(4)	0.54

5 Conclusions

We have presented a comparison of several tolerance and t-norm equations in prediction results of different treatments in Parkinson patients using fuzzy rough set theory (FRST). We have performed similar calculations for the symptoms predictions also using FRST. Our results demonstrated that attribute related to the eye movements are important and gave better predictions than only classical neurological measurements. This work is continuation of our previous papers where rough set theory (RST) was used. The global coverage results were better when FRST was used, however the global accuracy was higher with RST, but number of measurement is relatively small. A big advantage of the eye movement (EM) measurements is that they might be perform without doctor help, objectively with high precision and in the near future at patient's home. With help of the data mining methods such as RST or FRST these data can be automatically evaluated in order to give instant, objective advice to individual patient – it is the future method related to the tele-medicine. However, in order to be able to use the analyzed methods in practical applications, we need to perform measurements and confirm our results on larger group of patients that is actually in work in-progress.

References

1. Siegelmann, H.T.: Neural and super-turing computing. Mind. Mach. **13**, 103–114 (2003)
2. Przybyszewski, A.W., Linsay, P.S., Gaudiano, P., Wilson, C.M.: Basic difference between brain and computer: integration of asynchronous processes implemented as hardware model of the retina. IEEE Trans Neural Netw. **18**, 70–85 (2007)
3. Przybyszewski, A.W., Szlufik, S., Dutkiewicz, J., Habela, P., Koziorowski, D.M.: Machine learning on the video basis of slow pursuit eye movements can predict symptom development in Parkinson's patients. In: Nguyen, N.T., Trawiński, B., Kosala, R. (eds.) ACIIDS 2015. LNCS, vol. 9012, pp. 268–276. Springer, Heidelberg (2015)
4. Przybyszewski, A.W., Kon, M., Szlufik, S., Dutkiewicz, J., Habela, P., Koziorowski, D.M.: Data mining and machine learning on the basis from reflexive eye movements can predict symptom development in individual Parkinson's patients. In: Gelbukh, A., Espinoza, F.C., Galicia-Haro, S.N. (eds.) MICAI 2014, Part II. LNCS, vol. 8857, pp. 499–509. Springer, Heidelberg (2014)
5. Hikosaka, O., Takikawa, Y., Kawagoe, R.: Role of the basal ganglia in the control of purposive saccadic eye movements. Physiol. Rev. **80**, 953–978 (2000)
6. Jones, G.M., DeJong, J.D.: Dynamic characteristics of saccadic eye movements in Parkinson's disease. Exp. Neurol. **31**, 17–31 (1971)
7. Ladda, J., Valkovič, P., Eggert, T., Straube, A.: Parkinsonian patients show impaired predictive smooth pursuit. J. Neurol. **255**, 1071–1078 (2008)
8. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht (1991)
9. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets Syst. **126**, 137–156 (2002)
10. Riza, L.S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Ślęzak, D., Benítez, J.M.: Implementing algorithms of rough set theory and fuzzy rough set theory in the R package, "RoughSets". Inf. Sci. **287**, 68–69 (2014)
11. Hu, Q., Yu, D., Pedrycz, W., Chen, D.: Kernalized fuzzy rough sets and their applications. IEEE Trans. Knowl. Data Eng. **23**, 1471–1649 (2011)
12. Cornelis, C., De Cock, M., Radzikowska, A.: Fuzzy rough sets: from theory into practice. In: Pedrycz, W., Skowron, A., Kreinovich, V. (eds.) Handbook of Gradual Computing, pp. 533–552. Wiley, Chichester (2008)
13. Jensen, R., Cornelis, C., Shen, Q.: Hybrid fuzzy-rough rule induction and feature selection. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1151–1156 (2009)
14. Jensen, R., Shen, Q.: New approaches for fuzzy-rough feature selection. In: Proceedings of the 19th International Conference on Fuzzy Systems, vol. 149, issue no. 1, pp. 5–20 (2005)
15. Bazan, J., Nguyen, H.S., Nguyen, T.T., Skowron, A., Stepaniuk, J.: Decision rules synthesis for object classification. In: Orłowska, E. (ed.) Incomplete Information: Rough Set Analysis, pp. 23–57. Physica, Verlag (1998)