J Intell Inf Syst
https://doi.org/10.1007/s10844-018-0502-y

Analysis of medications change in Parkinson's disease progression data

Anita Valmarska[1,2] · Dragana Miljkovic[1] · Nada Lavrač[1,2,3] · Marko Robnik-Šikonja[4]

Received: 17 July 2017 / Revised: 9 February 2018 / Accepted: 19 February 2018

Abstract Parkinson's disease is a neurodegenerative disorder that affects people worldwide. Careful management of patient's condition is crucial to ensure the patient's independence and quality of life. This is achieved by personalized treatment based on individual patient's symptoms and medical history. The aim of this study is to determine patient groups with similar disease progression patterns coupled with patterns of medications change that lead to the improvement or decline of patients' quality of life symptoms. To this end, this paper proposes a new methodology for clustering of short time series of patients' symptoms and prescribed medications data, and time sequence data analysis using skip-grams to monitor disease progression. The results demonstrate that motor and autonomic symptoms are the most informative for evaluating the quality of life of Parkinson's disease patients. We show that Parkinson's disease patients can be divided into clusters ordered in accordance with the severity of their symptoms. By following the evolution of symptoms for each patient separately, we were able to determine patterns of medications change which can lead to the improvement or worsening of the patients' quality of life.

✉ Anita Valmarska
anita.valmarska@ijs.si

Dragana Miljkovic
dragana.milkovic@ijs.si

Nada Lavrač
nada.lavrac@ijs.si

Marko Robnik-Šikonja
marko.robnik@fri.uni-lj.si

[1] Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

[2] Jožef Stefan International Postgraduate School, Jamova 39, 1000 Ljubljana, Slovenia

[3] University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia

[4] Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia

Published online: 20 March 2018

Keywords Parkinson's disease · Quality of life indicators · Clustering · Short time series · Skip-grams

1 Introduction

Parkinson's disease is a neurodegenerative disorder that affects people worldwide. Due to the death of nigral neurons, there are changes in dopamine levels in the human brain causing several motor symptoms: tremor, rigidity, bradykinesia and postural instability. In addition to motor symptoms, Parkinson's disease is associated also with non-motor symptoms, which include cognitive, behavioral, and autonomic problems. These symptoms significantly decrease the quality of life of the patients affected by Parkinson's disease.

Over 6.3 million people have the condition worldwide (European Parkinson's Disease Association 2016). In Europe, more than one million people live with Parkinson's disease and this number is expected to double by 2030 (Dorsey et al. 2007). Parkinson's disease is the second most common neurodegenerative disease (after Alzheimer's disease) and its prevalence continues to grow as the population ages. Currently, there is no cure for Parkinson's disease. The reasons for the cell death are still poorly understood. The management of symptoms is of crucial importance for patients' quality of life, mainly addressed with antiparkinson medication, such as levodopa and dopamine agonists.

While numerous studies address specific aspects of the disease, there are few research efforts that adopt a holistic approach to disease management (Gatsios et al. 2016). The PERFORM (Tzallas et al. 2014), REMPARK (Samà et al. 2015) and SENSE-PARK (SENSE-PARK 2016) systems are intelligent closed-loop systems that seamlessly integrate a range of wearable sensors (mainly accelerometers and gyroscopes), constantly monitoring several motor signals of the patients and enabling the prescribing clinicians to remotely assess the status of the patients, given a real-time image of each patient's condition. Based on individual patient's response to the prescribed therapy (manifested by the change of the motor symptoms), the physician is able to adjust medication schedules and personalize the treatment (Gatsios et al. 2016). However, no data mining paradigms are used in the mentioned systems.

In the development of the PD_manager's m-Health platform for patient-centric Parkinson's disease management (PD_manager: m-Health platform for Parkinson's disease management 2015), one of the investigated approaches is data mining, aiming to provide decision support to clinicians and patients in personalized disease management. The individual patient's data, recorded in consecutive visits to the prescribing physician, are collected from different sources offering different 'views' of the data describing the same patient by multiple distinct feature sets. This setting suggests a multi-view learning approach.

Multi-view learning—a relatively new but well-established machine learning technique—is often appropriate for this type of data, as it aims to build models from multiple views (multiple data sets) by considering the diversity of different views (Xu et al. 2013). These views represent data obtained from multiple sources or different feature subsets and describe the same set of examples. We decided for a multi-view clustering approach, aiming to construct disjoint partitioning of objects (patients) described by multiple feature sets. This partitioning is aimed at identifying clusters of patients that share similar symptoms which enables automatic detection of interesting patterns.

Our work explores and tries to give answers to important medical questions which nobody (to the best of our knowledge) has tried to answer: How medications therapy of Parkinson's disease patients changes in response to the patients' change of overall status, and what are the directions in which the disease would develop based on the patients' symptoms and

their therapies. The goal of this paper is to develop a new clustering-based methodology for disease progression data, which will—based on the patients' allocation to clusters at given time points and their history of medication therapies—be able to make suggestions about modifications of particular patient's therapy, with the aim to improve the patient's quality of life. Patients' allocation to clusters represent their disease status, and their changed cluster allocation through time represents their disease progression. The analysis of the clusters can reveal what is the most common status of the patients, and the analysis of cluster changes can reveal how their symptoms change in time. Learning on the history of changes between clusters allows us to infer significant features and relevant medications changes for groups of patients and to suggest medications changes for the individual patients.

In order to increase the robustness of our results, we model the sequences of changes of patient's status between the clusters by using *skip-grams* (Guthrie et al. 2006), an approach upgrading the more standard n-grams approach (Broder et al. 1997) that is regularly used in the analysis of data sequences. The introduction of skip-grams results in increased number of investigated n-grams, providing a more stable distribution of the possible cluster changes.

This paper significantly extends our previous work (Valmarska et al. 2016). We extended the methodology for analysis of Parkinson's disease data to include three threads of clustering (Section 4). A pseudo code of the approach for dividing Parkinson's disease patients into groups with similar symptoms and ordering these groups of patients in accordance with the severity of their overall status is outlined in Section 4.2. The changes of patients antiparkinson medications dosages in relation to the change of their overall status is explored in Section 4.3, where we introduce Algorithm 2 to determine the change in medications dosage with respect to the change of patient's status. In Section 4.4 we present the skip-grams based approach for determining groups of patients with different patterns of disease progression based on the changes of their overall status. Finally, we have significantly extended the empirical evaluation in Section 5 by updating the previous symptoms analysis and medications analysis results with the results for determining the number of clusters and patterns of disease progression. We also present the results of detailed analysis of patients who were identified as following a certain pattern of disease progression.

The paper is structured as follows. After presenting the motivation, the background and the related work in Section 2, Section 3 describes the Parkinson's Progression Markers Initiative (PPMI) data (Marek et al. 2011) used in our experiments. Section 4 proposes the methodology for analyzing the Parkinson's disease data through clustering of short time series symptoms data and connecting the changes of symptoms-based clustering of patients to the changes in medication therapies with the goal to find treatment recommendation patterns and disease progression patterns. The latter is addressed by introducing the so-called skip-grams for analyzing the cluster change patterns and the progression of the disease. Section 5 presents the results of data analysis, tested on two data set variants. Finally, Section 6 presents the conclusions and ideas for further work. The paper contains four appendices which contain detailed results of analyzes: comparison of clustering algorithms (Appendix A), unsupervised feature selection (Appendix B), evaluation of different views in multi-view clustering (Appendix C), and descriptive rules for multi-view clusters (Appendix D).

2 Background

Parkinson's disease is a heterogeneous neurodegenerative condition with different clinical phenotypes, genetics, pathology, brain imaging characteristics and disease duration (Foltynie et al. 2002). This variability indicates the existence of disease subtypes. Moreover,

Parkinson's disease symptoms overlap with symptoms from other diseases, thus hampering the diagnosis of new PD patients and decreasing the overall success of the diagnosis process. Only 75% of clinical diagnoses of Parkinson's disease are confirmed to be idiopathic Parkinson's disease at autopsy (Hughes et al. 1992).

Given the heterogeneous nature of Parkinson's disease (PD), the nature of data describing PD patients is also heterogeneous, possibly gathered in different databases. Our data set (Marek et al. 2011) contains symptoms of patients suffering from Parkinson's disease where the symptoms are divided into several views. We test the union of all views with standard clustering approaches as well as several subsets of views using multi-view clustering in order to identify clusters of patients that share similar symptoms.

Patients' symptoms change through time depending on the received therapies, development of the disease, everyday habits, etc. We treat patients' symptoms at each time point as one training instance. This leads to patients' allocation to different clusters in different time points depending on the progression of the disease. We aim to suggest modifications of the medication treatments based on identified migration patterns of patients from one cluster to another with the goal to keep the patients in the clusters with symptoms that allow a good quality of life. To reach this goal we developed a new clustering-based methodology for disease progression data.

The reminder of this section presents Parkinson's disease related data mining research, an overview of relevant multi-view clustering approaches, and a short overview of methods for short time series analysis, including the introduction of skip-grams for sequence data analysis.

2.1 Parkinson's disease related data mining research

Data mining research in the field of Parkinson's disease (PD) can be divided into four groups: classification of PD patients, detection of PD symptoms, detection of subtypes of PD patients, and assessing success of deep brain stimulation surgery as a last resort in the treatment of Parkinson's disease patients.

The use of classification techniques offers decision support to specialists by increasing the accuracy and reliability of diagnosis and reducing possible errors. Gil and Johnson (2009) use Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to distinguish PD patients from healthy subjects. Ramani and Sivagami (2011) compare the effectiveness of different data mining algorithms in the diagnosis of PD patients.

Tremor is one of the symptoms strongly associated with Parkinson's disease. Several methods for numerical assessment of the intensity of tremor have been proposed. These methods include time series analysis (Timmer et al. 1993), spectral analysis (Riviere et al. 1997) and non-linear analysis (Riviere et al. 1997) and they address tremor detection and quantification. Recent works are based on body fixed sensors (BFS) for long-term monitoring of patients (Patel et al. 2009).

In the course of their disease, patients are prescribed antiparkinson medications therapies in order to control the troubling symptoms. As the disease progresses, the medications treatment can become ineffective and—as a last resort—clinicians use deep brain stimulation (DBS) surgery to control the Parkinson's disease symptoms. Data mining research confirms that DBS significantly improves the patients' motor function (Liu et al. 2014). Depending on the chosen method for DBS, a great reduction in dose of medication, or conservation of cognitive functions can be achieved. In order to predict the neurological effects related to different electrode-contact stimulation, Szymański et al. (2015) tracked the connections between the stimulated part of subthalamic nucleus and the cortex with the help of diffusion tensor imaging (DTI).

Identification of Parkinson's disease subtypes is presented in the work of Lewis et al. (2005), and has been confirmed by the conclusions from Reijnders et al. (2009) and Ma et al. (2015). While clustering usually focuses on patient grouping with the aim of diagnosing new patients, none of the listed methods follows the progression of the disease, and to the best of our knowledge, no data mining research in the field of Parkinson's disease analyzed the development of the disease in combination with the medications that the patients receive. Identification of groups of patients based on the similarity of their symptoms and the clinicians' reaction with medications modification in order to keep the patients as stable and in good status as possible, can be helpful in the assignment of personalized therapies and an adequate patient treatment. For that purpose, we propose a methodology for identification of groups of patients based on the severity of their symptoms, determination of disease progression, and the consequent patterns of medications modifications.

2.2 Multi-view clustering

Multi-view clustering is concerned with clustering of data by considering the information shared by each of the separate views. Many multi-view clustering algorithms initially transform the available views into one common subspace (early integration), where they perform the clustering process (Xu et al. 2013). Chaudhuri et al. (2009) propose a method for multi-view clustering where the translation to a lower vector space is done by using Canonical Correlation Analysis (CCA). Tzortzis and Likas (2009) propose a multi-view convex mixture model that locates clusters' representatives (exemplars) using all views simultaneously. These exemplars are identified by defining a convex mixture model distribution for each view. Cleuziou et al. (2009) present a method where in each view they obtain a specific organization using fuzzy k-means (Bezdek 1981) and introduce a penalty term in order to reduce the disagreement between organizations in the different views. Cai et al. (2013) propose a multi-view k-means clustering algorithm for big data. The algorithm utilizes a common cluster indicator in order to establish common patterns across the views.

Co-training (Blum and Mitchell 1998) is one of the earliest representatives of multi-view learning. This approach considers two views consisted of both labeled and unlabeled data. Using labeled data, co-training constructs a separate classifier for each view. The most confident predictions of each classifier on the unlabeled data are then used to iteratively construct additional labeled training data. Kumar and III (2011) apply the co-training principle (Blum and Mitchell 1998) in unsupervised learning. Clustering is performed on both views, then cluster points from one view are used to modify the clustering structure of the other view. Appice and Malerba (2016) employ the co-training principle in the multi-view setting for process mining clustering. The above-mentioned approaches presume that each of the respective views is capable of producing clusters of similar quality when considered separately. He et al. (2014) do not make that presumption. They combine multiple views under a principled framework, CoNMF (Co-regularized Non-negative Matrix Factorization), which extends NMF (Non-negative matrix factorization) for multi-view clustering by jointly factorizing the multiple matrices through co-regularization. The matrix factorization process is constrained by maximizing the correlation between pairs of views, thus utilizing information from each of the considered views. CoNMF is a multi-view clustering approach with intermediate integration of views, where different views are fused during the clustering process. The co-regularization of each pair of views makes the clustering process more robust to noisy views. The decision to use the CoNMF approach in our work was made based on this algorithm property and on the availability of its Python code.

2.3 Analysis of short time series

A time series is a series of data points indexed in time order. Time series data analysis was used to study a wide range of biological and ecological systems (Bence 1995). The use of time series allows for studying the dynamics of a system. Short time series (8 points or less) constitute more than 80% of all time series data sets (Ernst et al. 2005). The small number of available time points does not allow for identification of statistically significant temporal profiles (Ernst and Bar-Joseph 2006). Bence (1995) examines methods for adjusting confidence intervals of the mean and parameters of a linear regression for autocorrelation. De Alba et al. (2007) suggest that simpler models can be more effective on short time series. They show that the Bayesian approach is superior to the traditional approach when applied to short time series but inferior when applied on longer time series (De Alba et al. 2007). Most of the research in short time series analysis is related to the analysis of short time series microarray gene expression data. Ernst et al. (2005) present a method for clustering of short time series gene expression data, followed by the introduction of the STEM (Short Time-series Expression Miner) software program (Ernst and Bar-Joseph 2006) specifically designed for the analysis of short time series microarray gene expression data.

In the healthcare domain, Choi et al. (2017) incorporate temporal modeling using the recurrent neural network (RNN) model to predict heart failure. Imhoff et al. (1998) apply short time series analysis to monitor lab variables after liver surgery, and to offer support to clinicians in their decision-making process for the treatment of acute respiratory distress syndrome. Schieb et al. (2013) evaluate the clustering of stroke hospitalization rates, patterns of the clustering over time, and associations with community level characteristics. They generate clusters of high and low-stroke hospitalization rates during two periods of time. According to the place of residence of patients, counties in USA are assigned to a cluster. Following the transition of counties between clusters between these two periods, counties are labeled as having a persistently high, transitional, or persistently low-stroke hospitalization rate.

Murugesan et al. (2017) present a hierarchical multi-scale approach for visualizing spatial and functional cluster evaluation patterns. Their visualization method is two-stage method based on sequence of community detection at each time stamp and community tracking between steps. Greene et al. (2010) address the issue of identifying communities in dynamic networks. Appice (2017) uses social network analysis as a basic approach for organizational mining, aimed at understanding the life cycle of a dynamic organizational structures.

Zhao et al. (2017) explore different representations of temporal data from electronic health records to improve prediction of adverse drug events. They obtain sequences of symbols by transforming time series of individual feature into strings (Lin et al. 2007). These strings reflect the temporal nature of the original values. Results from their empirical investigation show that transformation of sequences to tabular form based on edit distance of sub-sequences to representative shaplets leads to improvements in the predictive performance. This approach reduces the feature sequence diversity by finding informative random sub-sequences. The goal of Zhao et al. (2017) is to predict whether patients will develop adverse drug reactions. They use the history of patients symptoms in order to predict a single event (adverse drug event: yes or no), while we follow the patients' disease development and changes in their overall status as a result of therapy changes. Another difference is our use of skip-grams which reduces noise and enforces strong transition patterns.

To the best of our knowledge, the temporal nature of medical data has not been explored in research directed toward determining the progression of a particular disease and determining the therapy recommendations in order to stabilize the disease progression. We present a clustering based methodology on short time series symptoms data of Parkinson's disease

patients in an attempt to discover how the disease develops through time, reflected by the change of patients' symptoms. Simultaneously, we use the temporal data about their medications therapy to determine how clinicians react to patients' symptoms changes. Each Parkinson's disease patient is described with his/her symptoms and medications treatment through time. The temporal data is flattened to records from single time points, referred in this manuscript as instances, where any change of patients' symptoms between two consecutive points is referred as change in their status. Changes in status are then connected to possible changes in medications therapies.

2.4 Skip-grams for sequence data analysis

Patient's allocation to clusters in sequential time points can be viewed as a sequence of items. Analysis of contiguous sequences of items for every patient's cluster allocation can provide an insight into the disease progression and reveal patterns how (and how often) the patient's symptoms improve or degrade.

In this paper we use an approach to sequence data analysis, where we borrow the methodology initially developed in the field of natural language processing (NLP). In NLP, a contiguous sequence of n items from a given sequence of text or speech is called an n-gram (Broder et al. 1997). Skip-grams are a generalization of n-grams in which the components (typically words) need not be consecutive in the text under consideration but may leave gaps that are skipped over (Guthrie et al. 2006). They provide a way of overcoming the data sparsity problem found with conventional n-gram analysis.

Another use of skip-grams is in producing word embeddings into a vector form to reduce dimensionality and sparsity of bag-of-words representation. Mikolov et al. (2013) proposed word2vec embedding based on deep learning, which has subsequently been used in many NLP applications, including some with clinical text data (Minarro-Giménez et al. 2013; De Vine et al. 2014) (PubMed abstracts, disease progression reports) and to learn relationships between clinical processes or unified medical language system (UMLS) concepts (Choi et al. 2017). Our use of skip-grams is entirely different as we do not use embeddings but use skip-grams directly as a more robust version of n-grams.

In the context of our analysis, skip-grams allow for robust identification of frequent paths through clusters and reveal typical disease progression patterns. The patient's overall status at a given visit to the clinician, as determined by the (patient, visit) pair cluster assignment, can be seen as an item, and changes of clusters as sequences of items, which can be analyzed with the skip-grams based approach developed in NLP. This is novel in the analysis of Parkinson's disease data and allows us to follow the progression of the patient's overall status without taking into account noise in the form of sudden changes in the patient's status. Such changes are not necessary due to Parkinson's disease, but can be attributed to other stressful events in the patient's life (such as loss of a pet, loss of a loved one, etc). To the best of our knowledge, there has not been any study involving skip-grams that uses the actual symptoms of patients in order to explore patient's disease progression and the clinicians' response by changing the medications therapy. A formal definition of skip-grams and their use are presented in Section 4.4.

3 Data

In this study, we use the PPMI data collection (Marek et al. 2011) gathered in the observational clinical study to verify progression markers in Parkinson's disease. The PPMI data

collection consists of data sets describing different aspects of the patients' daily life. Below we describe the selection of PPMI data used in the experiments.

3.1 PPMI symptoms data sets

The medical condition and the quality of life of a patient suffering from Parkinson's disease can be determined using the Movement Disorder Society (MDS) sponsored revision of Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (Goetz et al. 2008). It is a questionnaire consisting of 65 questions concerning the progression of disease symptoms. MDS-UPDRS is divided into four parts. Part I consists of questions about the 'non-motor experiences of daily living'. These questions address complex behaviors, such as hallucinations, depression, apathy, etc., and patient's experiences of daily living, such as sleeping problems, daytime sleepiness, urinary problems, etc. Part II expresses 'motor experiences of daily living'. This part of the questionnaire examines whether the patient experiences speech problems, the need for assistance with the daily routines such as eating or dressing, etc. Part III is referred to as the 'motor examination', while Part IV concerns 'motor complications', which are mostly developed when the main antiparkinson drug levodopa is used for a longer time period. Each question is anchored with five responses that are linked to commonly accepted clinical terms: 0 = normal (patient's condition is normal, symptom is not present), 1 = slight (symptom is present and has a slight influence on the patient's quality of life), 2 = mild, 3 = moderate, and 4 = severe (symptom is present and severely affects the normal and independent functioning of the patient, i.e. her quality of life is significantly decreased).

Montreal Cognitive Assessment (MoCA) (Dalrymple-Alford et al. 2010) is a rapid screening instrument for mild cognitive dysfunction. It is a 30 point questionnaire consisting of 11 questions, designed to assess different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation.

Scales for Outcomes in Parkinson's disease – Autonomic (SCOPA-AUT) is a specific scale to assess autonomic dysfunction in Parkinson's disease patients (Visser et al. 2004). Physical Activity Scale for the Elderly (PASE) (Washburn et al. 1993) is a questionnaire which is a practical and widely used approach for physical activity assessment in epidemiologic investigations. The above data sets are periodically updated to allow the clinicians to monitor patients' disease development through time. Answers to the questions from each questionnaire form the vectors of attribute values.

Table 1 summarizes the symptoms data sets considered in our research. It lists the number of considered questions from each questionnaire, the range of attribute values, and the nature of the attribute values. All of the considered questions have ordered values, and—with the exception of questions from MoCA and PASE—increased values suggest higher symptom severity and decreased quality of life.

When considering the possibility of using a multi-view framework, the independence of the separate views should be inspected. In their work, Goetz et al. (2008) present that the MDS-UPDRS shows high internal consistency (Cronbach's alpha = 0.79—0.93 across parts). MDS-UPDRS across-part correlations range from 0.22 to 0.66. Reliable factor structures for each part are obtained (comparative fit index > 0.90 for each part), which supports the use of sum scores for each part, when compared to using a total score of all parts.

3.2 PPMI concomitant medications log

The PPMI data collection offers information about all of the concomitant medications that the patients used during their involvement in the study. These medications are described

Table 1 Characteristics of the questionnaire data used in the analysis

Questionnaire	Number of questions	Answers value range	Ordered values	Higher value indicates higher symptom severity
MDS-UPDRS Part I	6	0–4	Yes	Yes
MDS-UPDRS Part Ip	7	0–4	Yes	Yes
MDS-UPDRS Part II	13	0–4	Yes	Yes
MDS-UPDRS Part III	35	0–4	Yes	Yes
MoCA	11	0–1	Yes	No
PASE	7	1–2	Yes	No
SCOPA-AUT	21	0–3	Yes	Yes

by their name, the medical condition they are prescribed for, as well as the time when the patient started and (if) ended the medications therapy. For the purpose of our research, we initially concentrate only on whether the patient receives a therapy with antiparkinson medications, and which combination of antiparkinson medications the patient has received between each of the time points when the MDS-UPDRS test and the MoCA test were administered. The main families of drugs used for treating motor symptoms are levodopa, dopamine agonists and MAO-B inhibitors (National Collaborating Centre for Chronic Conditions 2006). Medications which treat Parkinson's disease-related symptoms but are not from the above-mentioned groups of medications are referred to as *other*.

3.3 Experimental data

Symptoms of patients suffering from Parkinson's disease are grouped into several data sets, representing distinct views of the data. These views consist of data from MoCA test, motor experiences of daily living, non-motor experiences of daily living, complex motor examination data, etc. For each patient these data are obtained and updated periodically (on each patient's visit to the clinician)—at the beginning of the patient's involvement in the PPMI study, and approximately every 6 months, in total duration of 5 years—providing the clinicians with the opportunity to follow the development of the disease. The visits of each patient can be viewed as time points, and the collected data on each visit is the data about the patient in the respective time point. All time points collected for one patient form a short time series.

In the experiments we address two settings: the analysis of *merged symptoms data* and the analysis of *multi-view symptoms data*.

Merged symptoms data are represented in a single data table, constructed by using the sums of values of attributes of the following data sets: MDS-UPDRS Part I (subpart 1 and subpart 2), Part II, Part III, MoCA, PASE, and SCOPA-AUT.[1] Goetz et al. (2015) use sums of symptoms values as an overall severity measure of a given aspect of Parkinson's disease. Similarly, we use sums of attribute values from different data sets to present the overall status of patients concerning respective aspects of their everyday living. Table 2 outlines the attributes used to construct the merged symptoms data, together with their range of values. This is a simplified representation using seven attributes, each representing the severity of symptoms of a given symptoms group, which proved to be valuable in the initial experiments (Valmarska et al. 2016).

[1] Appendix B presents the clustering quality results on data set obtained by feature selection.

Table 2 List of attributes used in the merged symptoms data set

Dataset	Attribute name	Value range	Higher value indicates higher symptom severity
MDS-UPDRS Part I	NP1SUM	0–24	Yes
MDS-UPDRS Part Ip	NP1PSUM	0–28	Yes
MDS-UPDRS Part II	NP2SUM	0–52	Yes
MDS-UPDRS Part III	NP3SUM	0–138	Yes
MoCA	MCATOT	0–30	No
PASE	PASESUM	0–24	No
SCOPA-AUT	SCAUSUM	0–63	Yes

Multi-view symptoms data consist of seven data sets: MDS-UPDRS Part I, Part Ip, Part II, Part III, MoCA, SCOPA-AUT, and PASE. Each of these data sets consists of values of attributes, which represent answers to the questions from a particular questionnaire. Similarly to Goetz et al. (2015), we added an additional attribute to each data set, which is the sum of values of attributes in the given data set (this equals the values of individual attributes used in the merged symptoms data).

The experimental data include symptoms and medications data of 405 Parkinson's disease patients from the PPMI study. Out of these 405 patients, 265 patients are male and 140 are female. The youngest patient was 33 years old at the beginning of the study (baseline visit), and the oldest patient was 84 years old. The average age of patients is 61.09 years. The experimental data contains from 1 to 5 visits to the clinician. The average number of recorded visits is 3.321. The experimental data consist of 1,345 patient's visits and each visit is considered a separate data instance, representing the basic building block of the methodology described in Section 4.

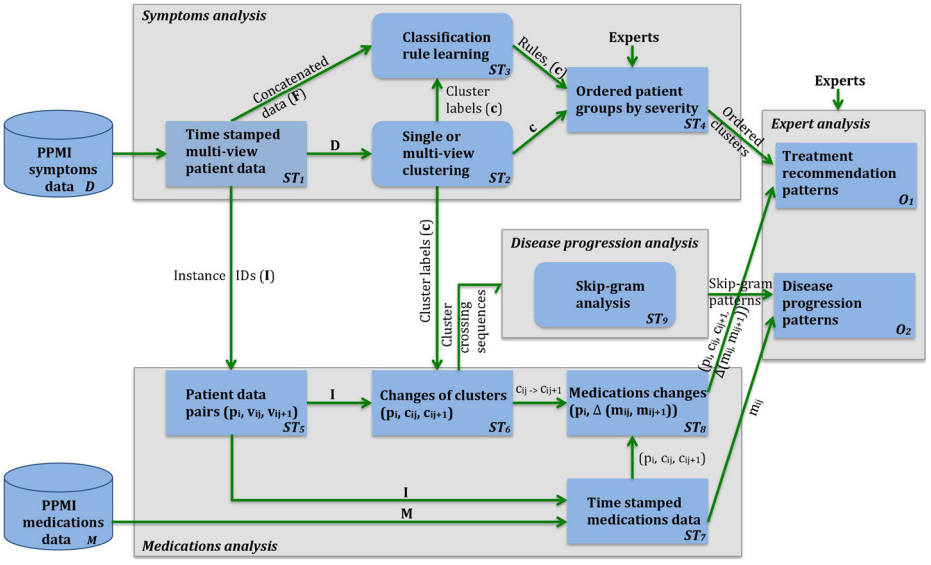


Fig. 1 Outline of the approach to Parkinson's disease quality of life data analysis

4 Methodology

To assist the clinicians in making decisions regarding the patients' therapy, we propose a procedure which involves a combination of clustering patients' symptoms data and the analysis of histories of patients' medication treatments, followed by disease progression analysis. Figure 1 shows an outline of the proposed methodology, which addresses changes of data over time (i.e. over several patient's visits) with the goal to suggest possible modifications of the medication treatment. Moreover, our goal is to analyze the rate of progression of Parkinson's disease and discover the most frequent patterns of symptoms change; we address this goal by using skip-grams on patients' changes of clusters. The usage of skip-grams can reveal groups of patients with an unusual pattern of symptoms change which deserve a more thorough look into the characteristics of that groups.

The input to the methodology are PPMI data sets of patient symptoms (described in Section 3.1) and the PPMI medications log data (described in Section 3.2), and the outputs are treatment recommendation patterns that can assist the clinician in deciding about further treatment of a patient, as well as the disease progression patterns providing insight into disease development. The methodology[2] consists of three separate threads whose outputs are combined to identify treatment recommendation patterns and disease progression patterns.

Symptoms analysis. The first thread, referred to as *Symptoms analysis* in the top part of Fig. 1, finds groups of patients with similar symptoms by grouping the instances, defined as (patient, visit) pairs. It uses clustering and describes the discovered patient groups with induced classification rules where classes correspond to individual cluster labels. Details of this thread are presented in Section 4.2.

Medications analysis. The second thread, referred to as *Medication analysis* in the bottom part of Fig. 1, is concerned with finding changes of medications and their dosages based on patients' symptoms changes between two consecutive visits to the clinician (e.g., disease aggravation, improvement or no change). In this thread we observe the patients moving from one cluster to another cluster in two consecutive time points, i.e. two consecutive visits to the clinician. The outcomes of the two threads are combined to a set of treatment recommendation patterns (i.e. increased/decreased/unchanged dosage of medications) for the four groups of medications mentioned in Section 3.2. We elaborate on this thread in Section 4.3.

Disease progression analysis. The third thread, referred to as *Disease progression analysis* in the middle part of Fig. 1, is concerned with finding patterns of disease progression, using skip-grams analysis on cluster crossing sequences. Details are given in Section 4.4.

The first step of the methodology is the construction of individual patient-visit pairs (p_i, v_{ij}), representing individual instances or items. For each patient p_i a set of pairs (p_i, v_{ij}) is constructed, where v_{ij} describes the symptoms recorded at an individual patient's visit to the clinician. These instances (patient-visit pairs) are the items representing the basic unit of analysis in the *Symptoms analysis* thread of the methodology. The attribute values of instance (p_i, v_{ij}) correspond to symptoms of patient p_i on visit j, and v_{ij} and v_{ij+1} correspond to two consecutive patient's visits. This is followed by clustering of instances.

The basic unit of the *Medications analysis* thread of the methodology are $(p_i, v_{ij}, c_{ij}, m_{ij}, v_{ij+1}, c_{ij+1}, m_{ij+1})$ tuples, where c_{ij} is the cluster label for instance (p_i, v_{ij}) and m_{ij} are the medications that patient p_i takes at the time of visit j. Elements c_{ij+1} and m_{ij+1} are

[2]The code is available upon request. Please note, we do not have a permission to share the data. Users can obtain permission from the Parkinson's Progression Markers Initiative (PPMI): http://www.ppmi-info.org/

the cluster label and prescribed medications of the same patient on visit $j+1$, i.e. at the time of the next visit.

The basic unit of the *Disease progression analysis* thread are patients' sequences of cluster crossings. Patient p_i cluster crossing sequence is Seq_i, defined as a sequence of cluster assignments for patient p_i at time points $v_{i1}, v_{i2}, ..., v_{ik_i}$, where v_{ij} correspond to the symptoms recorded at visit v_{ij} of patient p_i, and k_i is the number of visits to the clinician by patient p_i.

As our methodology is based on clustering, in Section 4.1 we first present cluster validity indices used to asses the quality of clusters produced by different tested methods.

4.1 Cluster validity indices

The number of groups (clusters) of similar patients was unknown before the start of the data analysis. In order to estimate the optimal number of clusters, we used internal cluster validity indices (Arbelaitz et al. 2013), which are—in the absence of ground truth labels—used to estimate the quality of generated clusters. The clustering quality is determined based on cluster *compactness*—how close are the related objects in each cluster, and cluster *separation*—how distinct or well-separated is each cluster from other clusters.

Many clustering validity indices (i.e. cluster quality measures) exist. We use three of the best performing indices from Arbelaitz et al. (2013): Silhouette analysis index (SA) (Rousseeuw 1987), Davies-Bouldin index (DB) (Davies and Bouldin 1979), and Calinski-Harabasz index (CH) (Caliński and Harabasz 1974). Below we present definitions and intuition behind these indices.

Let data set X be a set of N objects represented as vectors in an F-dimensional space, $X = \{x_1, x_2, ..., x_N\} \subseteq \Re^F$. Clustering of X is a set of disjoint clusters that partitions X into K groups. Clustering C is defined as disjoint partition of objects in X, $C = \{c_1, c_2, ..., c_K\}$, where $\bigcup_{c_k \in C} c_k = X$, $c_k \cap c_l = \emptyset$, $\forall k \neq l$. Centroid of a cluster c_k is defined as $\overline{c_k} = \frac{1}{|c_k|} \sum_{x_i \in c_k} x_i$. Similarly, the global centroid is defined as $\overline{X} = \frac{1}{N} \sum_{x_i \in X} x_i$. The Euclidean distance between two objects x_i and x_j is denoted as $d_e(x_i, x_j)$ (Arbelaitz et al. 2013).

Silhouette index is a normalized summation-type index. The compactness is measured based on the distance between all the objects in the same cluster and the separation is based on the nearest neighbor distance (Arbelaitz et al. 2013; Rousseeuw 1987; Kaufman and Rousseeuw 1990).

$$SA(C) = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{b(x_i, c_k) - a(x_i, c_k)}{\max\{a(x_i, c_k), b(x_i, c_k)\}} \tag{1}$$

where

$$a(x_i, c_k) = \frac{1}{|c_k|} \sum_{x_j \in c_k} d_e(x_i, x_j) \tag{2}$$

is the normalized distance of object x_i to all the objects in the same cluster (low values of this term are indicators of high compactness), and

$$b(x_i, c_k) = \min_{c_l \in C \setminus c_k} \left\{ \frac{1}{|c_l|} \sum_{x_j \in c_l} d_e(x_i, x_j) \right\} \tag{3}$$

is the normalized distance from object x_i to all objecs from its closest neighbor cluster (high values of this term are indicators of high separation). For each object, the quotient in (1) is a value between -1 and 1. A value close to 1 indicates that the object is well placed in its current cluster, while a value close to -1 indicates that it would be better placed in

the nearest cluster. Value 0 indicates a borderline quality of placement. An average over all objects gives an estimate on the overall quality of clusters. If there are too many or too few clusters as a result of inappropriate choice of the number of clusters K, many quotients will be low and the average score will reflect that.

Davies-Bouldin index estimates the compactness based on the distance of objects in a cluster to its centroid and the separation based on the distance between centroids (Arbelaitz et al. 2013; Davies and Bouldin 1979).

$$DB(C) = \frac{1}{K} \sum_{c_k \in C} \max_{c_l \in C \setminus c_k} \left\{ \frac{S(c_k) + S(c_l)}{d_e(\overline{c_k}, \overline{c_l})} \right\} \tag{4}$$

where

$$S(c_k) = \frac{1}{|c_k|} \sum\nolimits_{x_i \in c_k} d_e(x_i, \overline{c_k}) \tag{5}$$

is an average distance from objects in a cluster to its centroid. The $S(c_k)$ is a measure of the compactness for cluster c_k (the lower the value the more compact is the cluster). The quotients in (4) are indicators of separations between two clusters, c_k and c_l (the lower the quotient the better the two clusters are separated). By taking the maximum over these quotients we get the estimation of the worst case separation (i.e. for cluster c_k and its closest cluster). The average over these maxima is the value of DB index, whose lower values indicate better clusterings.

Calinski-Harabasz (CH) index estimates the compactness based on the distances from the objects in a cluster to its centroid (see the denominator below). The separation is based on the distance from the centroids to the global centroid $\overline{X}$ (see the nominator) (Arbelaitz et al. 2013; Caliński and Harabasz 1974).

$$CH(C) = \frac{N - K}{K - 1} \frac{\sum_{c_k \in C} d_e(\overline{c_k}, \overline{X})}{\sum_{c_k \in C} S(c_k)} \tag{6}$$

Factor $K - 1$ normalizes the distances of cluster centroids to global centroid, and factor $N - K = \sum_{k=1}^{K} (|c_k| - 1)$ normalizes the distances of objects to their centroids. Good clustering should have a large value in the nominator (large distances of clusters to global centroid) and a low value in the denominator (low distances of objects to their centroids) and therefore a large value of the CH score.

4.2 Symptoms analysis methodology

After constructing the instances—i.e. (patient, visit) pairs—in step $\mathbf{ST}_1$ of the methodology, the symptoms analysis thread (top of Fig. 1) consists of three further steps: clustering, rule learning and cluster ordering, corresponding to the individual steps of Algorithm 1 (lines 1-3).

The main input to Algorithm 1 is a set of symptoms views $\mathbf{D}$, describing the same n instances, which hold the symptoms data about p patients. This collection consists of m data sets (views). The k-th view ($1 \leq k \leq m$) is defined as $\mathbf{D}_k$, which is a matrix with n rows (the number of instances) and $|A_k|$ attributes. The concatenated data set, denoted as $\mathbf{F}$, is a matrix consisting of n rows and $\sum |A_k|$ columns (union of attributes across all the views). The medication data set, denoted by $\mathbf{M}$, is a matrix consisting of n rows and 4 columns—i.e. dosage data about the 4 PD medication groups. An auxiliary input is $\mathbf{I}$, a matrix which holds the indices of instances, defining the (p_i, v_{ij}) pairs.

The outputs of the algorithm are the assigned cluster labels $\mathbf{c}$ (vector of length n). The clustering of patients uses the provided views (symptoms data sets) and the chosen

clustering method (line 1 and line 2). The probability scores of dosage change of Parkinson's disease medications when patients' statuses improve or degrade are computed in line 4 with Algorithm 2) and are part of the medications change analysis methodology. These probability scores are estimates of medication impact on the change of symptoms.

Algorithm 1 PD_manager medications change methodology

Input : $\mathbf{D}=\{\mathbf{D}_1, ..., \mathbf{D}_m\}$ – collection of views, describing the instances;
m – number of data sets (views);
n – number of instances;
p – number of patients;
$|A_k|$ – number of attributes in k-th data set, $1 \leq k \leq m$;
$\mathbf{D}_k$ – a single view: matrix with n rows and $|A_k|$ columns;
$\mathbf{F}$ – concatenated data: matrix with n rows and $\sum |A_k|$ columns;
$\mathbf{A}=\{A_1 ... A_m\}$ – attribute space;
$\mathbf{M}$ – medications data set with 4 PD medications groups: matrix with n rows and 4 columns;
$\mathbf{I}$ – indices of patient-visit pairs;

Parameters: clusMethod – clustering method (*default*: if m=1 then k-means else CoNMF(pairwise));
descMethod – cluster description method (*default*: classification rule learning);

Output : $\mathbf{c}$ – assigned cluster labels (vector of size n);

```
// Form groups of similar instances.
1 if m=1 then c ← performSingleViewClustering(D,clusMethod);
2 else c ← performMultiViewClustering(D,clusMethod);

// Describe clusters and rank them according to quality of life
   indicators. Ranking is performed by the expert.
3 oClus ← describeClusters(F,c,descMethod)

// Obtain impact of medications.
4 medsChangeProb, medsChange ← getMedsChangeProbabilities(p,M,I,c)

// Obtain medications change patterns for cluster changes classified
   as positive or negative separately.
5 medsModificationPatterns ← getSummedMedsPatterns(medsChange,oClus)
6 visualize(medsModificationPatterns)
```

In step $\mathbf{ST}_2$ of the methodology outlined in Fig. 1, we perform clustering on instances i.e. patient's i symptoms recorded at a visit $v_{,ij}$ in order to determine groups of patients with similar symptoms. Note that the clustering step is performed once on the collection of views $\mathbf{D}$ which describes the instances. Our methodology can address both the merged symptoms data and the multi-view data analysis setting. The only difference is the clustering method applied in step $\mathbf{ST}_2$ of the methodology. In the case of merged symptoms data we performed k-means clustering (line 1 in Algorithm 1), while for clustering of the multi-view data we used the multi-view clustering approach proposed in He et al. (2014) (line 2 in Algorithm 1).

In the next step, $\mathbf{ST}_3$, we use the cluster labels ($\mathbf{c}$) as classes in rule learning in order to obtain meaningful descriptions of patients in each cluster (step $\mathbf{ST}_3$, line 3 in Algorithm 1).

Cluster labels obtained in step $\mathbf{ST}_2$ are input to step $\mathbf{ST}_3$ and are used as class labels in the rule learning process. The purpose of rule learning in step $\mathbf{ST}_3$ is to induce explanatory rules describing the induced clusters. These rules are presented to the experts (step $\mathbf{ST}_4$) to evaluate whether the induced clusters make sense and to determine an ordering of clusters according to the severity of symptoms of instances assigned to them. The rule sets describing the data are induced on a concatenated data set consisting of data sets considered in the clustering step ST_2 (input $\mathbf{F}$ of Algorithm 1).

The rule sets for each class variable are learned using our recently developed DoubleBeam-RL algorithm (Valmarska et al. 2017; Valmarska et al. 2015). This is a separate-and-conquer classification rule learning algorithm which uses two beams and separate heuristics for rule refinement and rule selection. Stecher et al. (2014) showed that the two phases of rule learning, rule refinement and rule selection, should be separated and use different rule evaluation heuristics in order to obtain rules with improved quality. They also introduce the idea of using the so-called *inverted heuristics* in the refinement phase in order to obtain rules that maximize the number of covered positive examples. By using the heuristics that take full advantage of the refinement and selection process separately, the DoubleBeam-RL algorithm is able to find rules which maximize the number of covered positive examples and minimize the number of covered negative examples, which is the goal of classification rule learning algorithms (Stecher et al. 2014). The DoubleBeam-RL algorithm generates rules with comparable accuracy to the rules generated by the state-of-the-art algorithms for classification rule learning (Valmarska et al. 2017), but as a side effect of using the inverted heuristics in the refinement phase, the induced rules have more conditions. The resulting longer rules with improved expressive power (Stecher et al. 2014; Michalski 1983) are preferred by the clinicians (Gamberger and Lavrač 2002). This is the reason for choosing the DoubleBeam-RL algorithm as the description tool in step $\boldsymbol{ST}_3$. Note that the DoubleBeam-RL algorithm does not perform rule pruning.

In the final step, $\mathbf{ST}_4$, the experts are presented with the descriptions of the obtained clusters, where the expert knowledge is used to interpret the obtained groups of patients and to order them according to the severity of symptoms exhibited by the patients assigned to them. The produced ordering of clusters may be total (all pairs of clusters are comparable) or partial (some clusters may not be comparable). Our methodology works for both cases as described below, but if in this step we get many incomparable clusters, this may be an indication that we have too many irrelevant or redundant attributes and we shall employ feature subset selection.

Based on the expert's interpretation of clusters and the ordering it produces, we take into account only comparable clusters and consider these cluster changes to be either positive or negative. When a patient moves from a cluster described by symptoms indicating worse quality of life to the one described by better quality of life indicators, we consider this change to be positive. A negative cluster change occurs when the symptoms of a patient degrade. Transitions between incomparable clusters are left out of our analysis.

In $\boldsymbol{O}_1$, we combine detected medications changes from step $\boldsymbol{ST}_8$ and cluster severity information from step $\mathbf{ST}_4$. The combined information contains medications changes for positive cluster changes and for negative cluster changes i.e. medications changes with improvement or aggravation of the patients' symptoms. Cluster changes are determined in line 5 of Algorithm 1 and the approach is further explained in Section 4.3.

4.3 Medications change analysis methodology

In this thread of the methodology (bottom of Fig. 1, lines 4–6 in Algorithm 1) we determine the medications changes that have occurred simultaneously with moves between clusters

observed in patients during two consecutive time points (two consecutive visits). An important benefit of our approach is that each patient provides a context (similar observed and unobserved variables) for himself/herself. By following the development of symptoms for each patient separately, we remove the influence of other conditions the patient is treated for.

The information about patients' assignment to clusters and their medication therapy in two consecutive visits is held in $(p_i, v_{ij}, c_{ij}, m_{ij}, v_{ij+1}, c_{ij+1}, m_{ij+1})$ tuples. In step $\boldsymbol{ST}_8$ on Fig. 1 we follow all patients through time. For each pair of patient's p_i consecutive visits to the clinician, we record the cluster change that has occurred between the two visits, $c_{ij} \rightarrow c_{ij+1}$, as well as the change in medications prescriptions, $\Delta(m_{ij}, m_{ij+1})$, which the patients received in the consecutive time points. For each antiparkinson drug group (levodopa, dopamine agonists, MAO-B inhibitors, and others) we record whether their dosage has increased, decreased or stayed unchanged between the two visits. Dosages of PD medications are translated into a common Levodopa Equivalent Daily Dosage (LEDD) which allows for comparison of different therapies (different medications with personalized daily plans of intake).

Algorithm 2 presents the pseudocode of the $getMedsChangeProbabilities$ function. It describes how we determine the changes of medications dosages co-occurring with shifts in patients' symptoms (characterized by a change of clusters). This function (called in line 4 of Algorithm 1) estimates the probability score of medications dosage changes when patients' symptoms have changed (patients have crossed clusters) or stayed the same (patient did not change clusters between two consecutive visits). Additionally, it also counts the type of medication modifications for each cluster crossing. The algorithm takes as an input patients' medications data **M**, the index data set **I**, and the assigned cluster labels **c**. The output are two matrices, $medsChangeProb$ and $medChange$ of the dimension $K \times K \times 4 \times 3$ (K is the number of clusters, we have 4 medication groups and 3 possible changes in severity of symptoms). Each cell of the output matrix $medsChangeProb$ contains a probability that a given medication group will change value (increase, decrease, or stay unchanged) for a certain cluster crossing. Similarly, the $medsChange$ matrix contains the number of changes of each type for each group and each crossing.

For each patient (line 5 in Algorithm 2), we track his/her status development through time. For each two consecutive visits (line 7), we register the clusters the patients were assigned to (lines 8 and 9). These consecutive cluster assignments represent a so-called *cluster crossing* (line 10). For each patient, we also follow therapy changes between two consecutive visits (lines 11 and 12). We consider therapy changes to be dosage changes of any of the antiparkinsonian medications (line 13). For each medications group, we record whether the LED dosage between two consecutive time has *increased*, *decreased*, or stayed *unchanged* (line 14). We record the number of therapy changes for each cluster crossing (line 15). The probability of medications change is calculated in line 24 of Algorithm 2 as the ratio between the recorded number of therapy modifications per cluster crossing and the number of cluster crossings. The output of Algorithm 2 are two matrices, $medsChangeProb$ and $medsChange$, described above.

Both matrices are returned to Algorithm 1. Matrix $medsChange$ is further processed in line 5 with function $getSummedMedsPatterns$. Based on the clusters ordered by the experts according to the severity of symptoms and the information on medications changes for each cluster crossing, we determine patterns of medications adaptations, related to the improvement or aggravation of patients' symptoms. Cluster crossings are classified as either positive or negative. We aggregate (sum) the medications change patterns from cluster changes of the same nature (positive or negative) to determine the patterns of medication modifications when the patients' status improved or worsened. The results are visualized in line 6 of Algorithm 1 (for the results, see Fig. 3).

Algorithm 2: The procedure estimates the probability of medications changes due to symptom changes

```
getMedsChangeProbabilities(p,M, I,c):
Input      : p – number of patients;
             M – patients' medications data;
             I – indices of patient-visit combinations;
             c – assigned cluster labels;
Parameters : K – number of clusters in c;
Output     : medsChangeProb;
             medsChange;

// Initialize number of cluster changes to 0. Number of clusters is K.
// noOfCrossings; matrix with K rows and K columns
noOfCrossings [1:K, 1:K] ← 0
// Initialize number of medications changes for each cluster change to 0.
// medsChangeNo; array of dimension K × K × 4 × 3 (4 medication groups, 3 changes).
medsChangeNo [1:K, 1:K, 1:4, 1:3] ← 0
// Initialize probabilities of medications changes for each cluster change to 0.
// medsChangeProb; array of dimension K × K × 4 × 3 (4 medication groups, 3 changes).
medsChangeProb [1:K, 1:K, 1:4, 1:3] ← 0
// For each patient check how cluster labels changed between consecutive time points.
for patient p_i in [1 : p] do
    // consecutive visits for a given patient
    patientsVisits ← I[p_i]
    for v_j, v_{j+1} in patientsVisits do
        // Patient's cluster assignement in two consecutive visits.
        prevCluster ← c[p_i][v_j]
        currCluster ← c[p_i][v_{j+1}]
        // Increase the number of crossings between prevCluster and currCluster.
        noOfCrossings[prevCluster,currCluster]+=1
        // Patient's medications therapy in two consecutive visits (time points).
        prevMeds ← M[p_i][v_j]
        currMeds ← M[p_i][v_{j+1}]
        // Compare medications therapy change between two consecutive visits.
           Patients can receive medications from four medications groups: levodopa,
           dopamine agonist, MAO-B, or other.
        for medsGroup in [1:4] do
            // dosage change can be either increased, decreased or unchanged
            change ← getChange(medsGroup, prevMeds, currMeds) medsChangeNo[prevCluster,
            currCluster,medsGroup, change]+=1
        end
    end
end
// Get medications change probability for each of increased, decreased or unchanged.
// Inspect all cluster crossings and medication groups.
for c_1 in [1:K] do
    for c_2 in [1:K] do
        clusterCrossings = noOfCrossings[c_1, c_2]
        for medsGroup in [1:4] do
            for change in [1:3] do
                medsChange = medsChangeNo[c_1, c_2,medsGroup,change]
                medsChangeProb[c1, c2,medsGroup, change] = medsChange / clusterCrossings
            end
        end
    end
end
return medsChangeProb, medsChange
```

4.4 Disease progression analysis using skip-grams

In this thread of the methodology (middle of Fig. 1, step $\boldsymbol{ST}_9$) we determine patterns of cluster changes, resulting in $\boldsymbol{O}_2$ combining the patients' cluster change patterns with the patients' medication data. This allows us to determine patterns of disease progression (indicated by the patterns of cluster change) and impact of medications on these patterns. Outputs $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are presented to the expert for analysis and validation.

Medical status of each patient's status at successive time points can be expressed as a sequence of clusters. The status of patient p_i at time point v_{ij} is responsible for patient's assignment to cluster c_{ij} at that point. Let Seq_i be a sequence of cluster assignments for patient p_i at time points $v_{i1}, v_{i2}, ..., v_{ik_i}$, denoted as $Seq_i = (c_{i1}, c_{i2}, ..., c_{ik_i}) \subseteq C$, where k_i is the number of visits to the clinician by patient p_i and $C = \{c_1, c_2, ..., c_K\}$ is the clustering (set of cluster labels) on the symptoms data. We denote the set of all cluster switching sequences for all patients as Seq

$$Seq = \bigcup_{1 \leq i \leq p} Seq_i, \tag{7}$$

where p denotes the number of analyzed Parkinson's disease patients.

The approach is inspired by natural language processing (NLP) approaches. In NLP, an n-gram is defined as a contiguous sequence of n items from a given sequence of text or speech. In order to analyze the patterns of symptoms changes across all the patients, we perform skip-gram analysis on Seq. A patient's sequence Seq_i can be regarded as an individual document in corpus Seq, where each cluster assignment c_{ij} represents the j-th word in document Seq_i.

The definition of k-skip-n-grams (Guthrie et al. 2006) for a document constructed from words $w_1 ... w_l$ can be expressed as

$$\{w_{i_1}, w_{i_2}, ..., w_{i_n} | \sum_{1 \leq j \leq n} i_j - i_{j-1} < k\} \tag{8}$$

Skip-grams reported for a certain skip distance k allow a total of k or less skips to construct the n-gram. Thus, 3-skip-n-gram results include 3 skips, 2 skips, 1 skip, and 0 skips. The 0-skip-n-grams are n-grams formed from adjacent words. The algorithmic construction of k-skip-n-grams starts with unigrams (which are 0-skip-1-grams) and progressively increases both the skip and the length of the sequence until the required k and n are reached.

We use skip-grams to determine the most frequent statuses of patients, and the most frequent patterns of their symptoms changes. Using skip-grams, the number of investigated n-grams significantly increases, thus providing more reliable introspection into cluster crossings. By skipping certain time points, we take into account that patients' statuses may occasionally result from other factors rather than the natural progression of the disease or the medication therapy. For example, the patient's non-motor symptoms (i.e. depression, apathy, etc.) may worsen due to a sudden death in the family, loss of a friend or loss of a pet. In other words, skip-grams make the resulting patterns more robust compared to the n-grams.

We present an example illustrating the advantage of using skip-grams instead of n-grams in analyzes of sequences. Lets say that a Parkinson's disease patient (p_i) has had 5 visits to the clinician. Based on the patient's symptoms, in each visit the patient was assigned to the following clusters $Seq_i = (1, 0, 2, 0, 1)$ (on visit 1, the patient was assigned to the cluster labeled as *1*, on the second visit, the patient was assigned to the cluster with label *0*, etc.).

The sets of sequences obtained for *bigrams*, *2-skip-bigrams*, *trigrams*, and *2-skip-trigrams* are presented below:

bigrams: {10, 02, 20, 01}
2-skip-bigrams: {10, 12, 10, 02, 00, 01, 20, 21, 01}
trigrams: {102, 020, 201}
2-skip-trigrams: {102, 100, 101, 120, 121, 101, 020, 021, 001, 201}

Using skip-grams to identify interesting patterns in short series of disease progression (reflected by cluster changes) is novel and we are not aware of other equally effective and noise-tolerant method for analysis of really short series. Another seemingly related approach would be to compute frequent itemsets used with association rules (Agrawal et al. 1993) but note that itemsets do not preserve temporal aspect of sequences which is an important information for disease progression.

5 Results of data analysis

The experimental work of this paper is divided into four parts. Initially, we are interested in whether Parkinson's disease patients can be divided into groups of patients with similar symptoms. After determining the appropriate number of clusters in Section 5.1, we report results of two experimental settings: i) using k-means clustering of the merged symptoms data set, and ii) using multi-view clustering on seven separate data sets (seven separate views). The results of both clustering experiments are presented in Section 5.2. This analysis was followed by an attempt to understand the effects of medications changes on the changes of patients' symptoms; these results are presented in Section 5.3. Finally, in Section 5.4 we present the results of experiments intended to find patterns in Parkinson's disease progression. The four groups of reported results were obtained with methodology described in Sections 4.1, 4.2, 4.3, and 4.4, respectively.

5.1 Determining the number of clusters

In order to determine the optimum number of clusters we ran the k-means clustering algorithm on the merged data set using different values for k. The obtained clusters were evaluated using the cluster validity indices introduced in Section 4.1. In terms of these cluster validity indices, better clustering quality is indicated by larger values of SA and CH indexes and lower values of DB index (Liu et al. 2010).

The results of k-means clustering presented in Table 3 show scores obtained for different values of k. The table indicates that k-means clustering produces the best clusters when the value of k is set to 2 or 3. The clustering quality decreases for $k > 3$, as indicated by all of the considered cluster validity indices.

We hypothesize that the reason for no difference in DB and CH indexes when $k = 2$ and $k = 3$, while there is a significant difference in SA, is due to differences how these indices are computed: DB and CH compare distances to centroids, while SA uses nearest neighbors between the instances.

Setting the value of k to 2 would divide patients into two groups: one with a good status and the other with a bad status of PD symptoms; this grouping would not take into account other values of symptoms except the ones characterized as either normal or very problematic for the patients. For this reason and to provide more variability we set the value of k to 3 to get three patient clusters instead of just two.

Table 3 Values of clustering validity scores for different number of clusters. Clusters are generated on the merged data set, using the *k*-means clustering algorithm

Number of clusters	Silhouette index (SA)	Davies-Bouldin index (DB)	Calinski-Harabasz index (CH)
2	0.516	0.916	162.540
3	0.347	0.916	162.540
4	0.371	1.266	54.103
5	0.279	1.133	40.545
6	0.276	1.458	32.412
7	0.258	1.489	26.992

Please note that the selection of k-means approach for clustering the merged data set was done after series of experiments. We checked three clustering approaches: k-means, k-medoids (Kaufman and Rousseeuw 1987), and DBSCAN (Ester et al. 1996). For each of the considered approaches, we evaluated the produced clusters using the validity indices SA (Rousseeuw 1987), DB (Davies and Bouldin 1979), and CH (Caliński and Harabasz 1974). Based on the results, we decided to use k-means as our clustering method of choice. Evaluation details for the considered clustering approaches can be found in Appendix A in Table 5.

5.2 Results of symptoms data analysis

To determine the progression of patients' symptoms, for each Parkinson's disease patient from our data set and for each two consecutive time points we investigated changes of clusters in which the patient participated. With the help of the expert, we order the clusters according to the quality of life indicators (i.e. severity of symptoms) of patients in the clusters. The evaluation of the quality of discovered clusters is two-fold. Clusters are initially evaluated using the internal cluster validity indices: SA, DB, and CH. The generated clusters are described by rules produced with the DoubleBeam-RL algorithm, and these descriptions are presented to experts. Based on the rules, experts order clusters according to the severity of symptoms of patients involved in each cluster.

5.2.1 Results of merged symptoms data analysis

The classification rules describing the clusters obtained from the merged symptoms data analysis are presented in Table 4. The rules indicate that the clusters are linearly ordered (with indexes 0, 1, and 2) and contain instances (patients symptoms recorded at a certain time point) with different severity of their motor symptoms. *Cluster 0* consists of instances with the sum of motor symptoms severity up to 22 (out of 138). Patients that have slightly worse motor symptoms are assigned to *Cluster 1* (sum of motor symptoms severity between 23 and 42). In *Cluster 2* there are patients whose motor symptoms significantly affect their motor functions (sum of motor symptoms severity greater than 42). The worsening of motor symptoms is followed by the aggravation of non-motor symptoms, mostly autonomic symptoms (sleeping, urinary, or constipation problems). This can be observed by the increased values of attributes SCAUSUM and NP2SUM in the rule sets describing *Cluster 1* and *Cluster 2*.

Inspection of the time line of cluster changes for a single patient. In order to illustrate the cluster changes for a patient, we subjectively chose a patient who already completed

Table 4 Rules describing clusters obtained by k-means clustering on the concatenated data set of attribute sums. Variables p and n denote the number of covered true positive and false positive examples respectively. We present the complete rules generated by the DoubleBeam-RL algorithm which does not prune its learned rules

Rule		p	n
Rules for cluster 0			
NP3SUM $\leq$ 20	$\rightarrow$ cluster = 0	488	4
NP3SUM $\leq$ 21 AND NP2SUM $\leq$ 6	$\rightarrow$ cluster = 0	321	0
NP3SUM = (19, 22] AND NP1SUM = 0	$\rightarrow$ cluster = 0	54	23
Rules for cluster 1			
NP3SUM = (22,30]	$\rightarrow$ cluster = 1	323	13
NP3SUM = (30, 39] AND SCAUSUM = (4, 10]	$\rightarrow$ cluster = 1	91	17
NP3SUM = (22, 42] AND NP2SUM = (0, 6]	$\rightarrow$ cluster = 1	206	6
NP3SUM = (22, 34] AND SCAUSUM = (10, 17] AND PASESUM >9	$\rightarrow$ cluster = 1	101	6
Rules for cluster 2			
NP3SUM >42	$\rightarrow$ cluster = 2	125	1
NP3SUM >37 AND NP1PSUM >5 AND MCAVFNUM $\leq$ 18	$\rightarrow$ cluster = 2	123	6
NP3SUM >30 AND NP2SUM >17	$\rightarrow$ cluster = 2	82	0
SCAUSUM >20 AND NP2SUM >9 AND MCAVFNUM $\leq$ 24	$\rightarrow$ cluster = 2	54	18
NP3SUM >30 AND SCAUSUM >11 AND NP2SUM >12	$\rightarrow$ cluster = 2	123	2
NP3SUM >36 AND SCAUSUM >6 AND NP2SUM >6 AND NP1PSUM >2	$\rightarrow$ cluster = 2	168	6

involvement in the PPMI study, and present her changes in the overall status in Fig. 2. The disease status can be tracked through changes in the patient's cluster assignments recorded during consecutive visits to the clinician. We also present the changes in medications therapy, made in order to keep the patient's symptoms as stable as possible.

We presented the figure (as well as the symptoms and medications data) to our consulting clinician for interpretation. He commented that the particular treatment was in accordance with the standard practice and guidelines for the treatment of Parkinson's disease patients. The usual practice is that clinicians almost always start with MAO-B inhibitors (such as Azilect) to protect neurons and later introduce dopamine agonists (such as Requip or Neupro) in order to manage Parkinson's disease (European Parkinson's Disease Association 2016). The usage of levodopa (Carbidopa/Levodopa) is delayed as long as possible—symptoms allowing—in order to avoid the side effects of prolonged usage of levodopa, such as dyskinesia and on/off fluctuations.

As evident from the diagram, the initial status of the patient was good. The clinician started the treatment of Parkinson's disease by introducing a MAO-B inhibitor (Azilect). Then clinician increased the dosage, trying to find an appropriate dosage for the specific patient. Once the patient's symptoms worsen (as indicated by the cluster changes between visits V04 and V06), the clinician introduced dopamine agonists to stabilize the symptoms. There were several adjustments aiming to find the appropriate dopamine agonist therapy

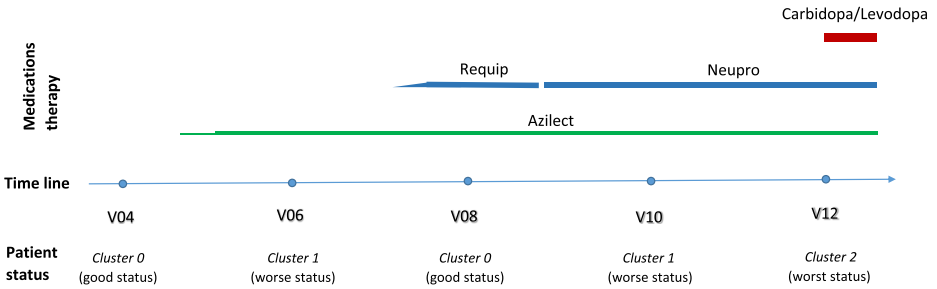


Fig. 2 Inspection of a cluster change time line of a single patient. Points on the time line present visits the patients has made to the clinician. Patient's medications therapy is presented by the groups of antiparkinson medications the patient has received during her involvement in the PPMI study. The color of medications therapy determines the group of antiparkinson medications—MAO-B inhibitors are presented with the green line, dopamine agonists are presented with the blue line, and levodopa based medications are presented with the red line on the top. Line width indicates the value of LEDD, i.e. the ticker the line the higher the value of LEDD. Endpoints of lines indicate beginnings and ends of treatments with particular medications. For example, the patient was treated with Requip (a dopamine agonist medication) starting between visit 6 (V06) and visit 8 (V08) and ending sometimes after visit 8. The treatment with Neupro (with almost the same value of LEDD) started immediately after the treatment with Requip stopped and was ongoing even after the patient finished her involvement in the PPMI study (V12)

for the patient: the clinician started with Requip and changed the medication's dosage several times (represented by the steep increase of the blue line). The patient initially reacted well to this change and her overall status was improved (V08). However, the status then worsened and the clinician changed the therapy by ending the intake of Requip and introducing Neupro. This medication change did not improve the patient's status at visit V10, and by visit V12 her status got even worse (our methodology assigned the patient to *cluster 2* at visit V12). Since the patient's status was bad and the quality of life has significantly declined, the clinician was forced to introduce levodopa.

5.2.2 Results of multi-view symptoms data analysis

In addition to analyzing the merged symptoms data, we performed a number of experiments on multi-view symptoms data consisting of seven separate symptoms data sets. In these experiments, we used the CoNMF multi-view clustering algorithm (He et al. 2014). Similarly to the merged view clustering approach, we tried to compare the clusters obtained by the multi-view approach by the severity of patients' symptoms assigned to them. The analysis has revealed that there were no significant intersections of the instances assigned to the clusters obtained by the multi-view approach compared to the clusters obtained by k-means ($k = 3$) clustering of the merged symptoms data set. Furthermore, given that the distinction between the three produced clusters was unclear, the ordering and comparison of the clusters was not possible This result means that we were not able to interpret the clustering produced by the CoNMF algorithm. In Appendix C we present the results of further analysis on impact different views have in the multi-view clustering process.

Results from Table 8,[3] and Table 10 in Appendix C show that the quality of clusters induced using the CoNMF approach is lower than the quality of clusters generated on

[3]Note that in Table 9 we present the Adjusted Random Index values where we compare the cluster similarity between the three best performing bi-view clustering settings.

the merged data set. Results reveal that it is beneficial to combine multiple data sets in order to obtain better clusters and better overall picture of the patients that were assigned to these clusters. However, when including new views, one must be careful, since the inclusion of seemingly uncorrelated views can hinder the performance of the multi-view approach. Results from Table 10 show that the best quality clusters are obtained when using only three data sets (views): SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III. Due to low quality of induced clusters, we decided not to investigate the changes of medications dosages with respect to the changes of clusters generated in the multi-view clustering setting. However, in future, we will consider also other multi-view clustering algorithms.

Rules discovered with the best multi-view clustering are presented in Tables 11, 12, and 13 in Appendix D. The groups of patients are described mostly by their motor symptoms and descriptions are supported by attributes from the data set SCOPA-AUT. The SCOPA-AUT data set contains information about the autonomic symptoms of patients, namely mostly constipation and urinary problems. Even though the resulting multi-view clustering is of low quality and the experts were not able to order the produced clusters by the severity of the symptoms of patients involved in them, the consulted experts were pleased with the discovery that autonomic symptoms from SCOPA-AUT play an important role in produced clusters, as recent research shows that autonomic symptoms can be a potential premotor marker of Parkinson's disease (Ceravolo et al. 2010).

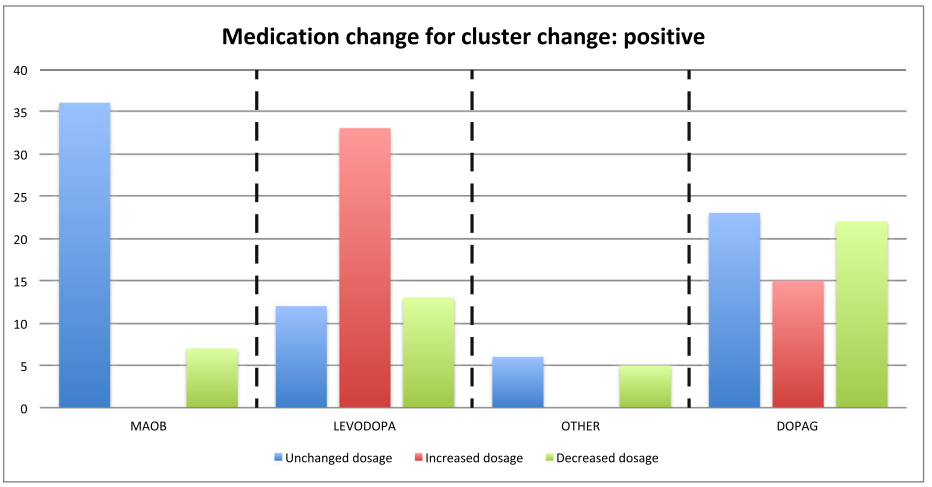
5.3 Results of medications change analysis

The experts were able to order clusters obtained from the merged symptoms data (presented in Table 4) by the severity of symptoms. The order was total (all clusters were comparable), so we assigned the three clusters indexes 0, 1, and 2 (lower index means lower severity of symptoms). When a patient moves from a cluster with a lower index to the one with a higher index, the patient's symptoms have worsened and we consider this change to be negative. A positive cluster change is recorded if the patient's symptoms have improved and the patient moves to a cluster with a lower index. The medications change patterns for positive and negative cluster change were obtained with the approach described in Section 4.3. The results are shown in Fig. 3.

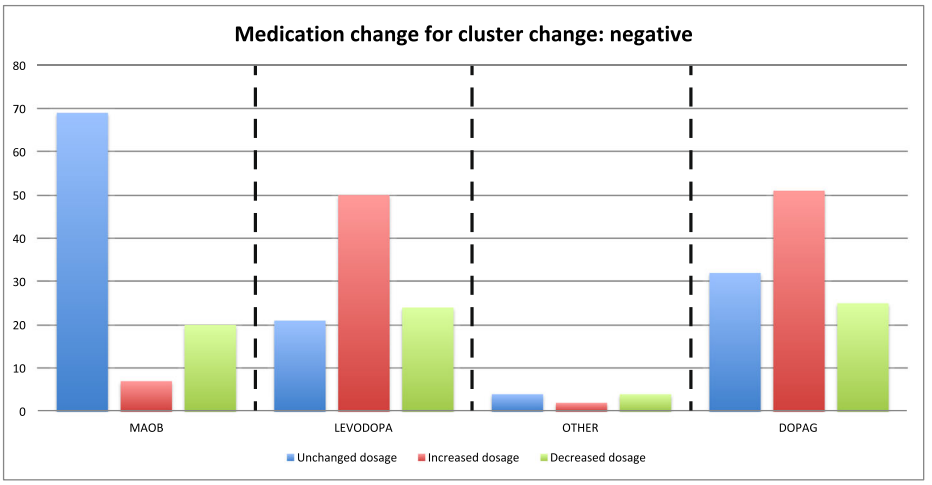
Figure 3a shows the medications changes when a positive cluster change has occurred. The red bars represent the number of times the dosage of medications from certain medication group has increased. Similarly, the number of times the medication dosage has decreased is shown in green. Blue bars present the number of times when a positive cluster change has occurred, but the medication dosage has stayed unchanged.

Figure 3b outlines the medications changes when a negative cluster change has taken place. These two graphs show patterns of medications modifications as a result of significant changes in the patient's status (patient's symptoms in two consecutive time points changed significantly, thus prompting a cluster change).

Figure 3 indicates that the patients' motor symptoms improve when the dosage of medications from the levodopa drug group is increased and the dosage of dopamine agonists is decreased or stays the same. When the dosage of both levodopa medications and dopamine agonists is increased the motor symptoms of the patients worsen. Clinicians prescribe and gradually increase the dosages of levodopa to handle the motor symptoms of patients. The usage of high dosages of dopamine agonists produces side effects affecting the non-motor



(a) Positive cluster change.



(b) Negative cluster change.

Fig. 3 Recorded Parkinson's disease medications changes when patient's cluster allocation has changed. Clusters were obtained from merged symptoms data set. A positive cluster change indicates that the patient's symptoms improved. A negative cluster change occurs when the patient's symptoms worsen. Medication groups are visually divided by vertical dashed lines

symptoms of the patients. A decrease of dosage eliminates these side effects and improves the patient's status.

5.4 Disease progression patterns

Figure 4 presents the results from the 3-skip-2-gram analysis of cluster crossings in the merged symptoms clustering setting. The results indicate that the patients' status is mostly stable over the considered time points. Patients tend to stay in the clusters they were initially assigned to. This is followed by a portion of patients whose symptoms worsen

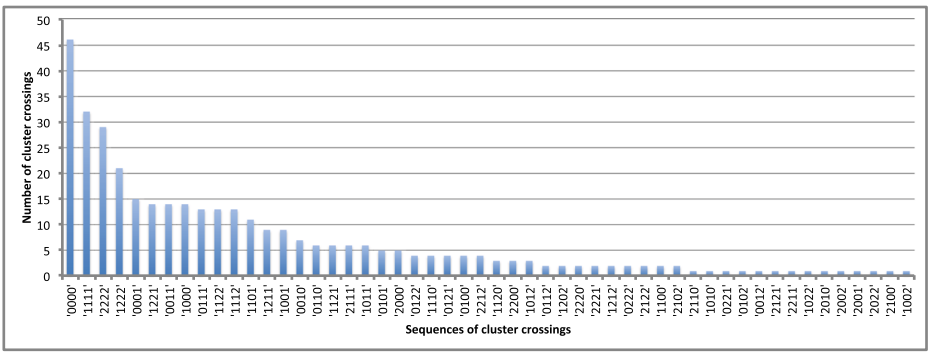


Fig. 4 Histogram resulting from 3-skip-2-gram analysis. The possible cluster crossings are listed on the X-axis (e.g., 01 indicates that a patient has moved from *Cluster 0* to *Cluster 1*), while the Y-axis represents the number of cluster crossings

(cluster crossings 01 and 12) and those whose symptoms improve (cluster crossings 10 and 21). These symptoms changes have all occurred gradually—patients have moved to the adjacent cluster. The number of patients whose symptoms have significantly changed (cluster crossings 02 and 20) is much lower.

The analysis of bigrams (2-grams) in Fig. 4 cannot reveal trends in patients' status over longer time period. Figure 5 presents the patterns of 4 almost consecutive cluster crossings obtained by 3-skip-4-gram analysis of the sequences of cluster crossings on the merged data set. It confirms the results from Fig. 4 which indicate that patients' status is usually stable and they tend to stay in the same cluster to which they were initially assigned.

Figure 5 reveals existence of interesting and slightly unexpected patterns of symptoms change: 1001, 0110, and 2000. We selected these sequences (subjectively) as patients' conditions are not steadily deteriorating and use them to illustrate our approach—the patients with similar symptoms have similar patterns of disease progression. We discuss groups of patients with 0110 and 2000 pattern below.

The analysis of patients with the 0110 cluster change pattern reveals that these are younger patients (50–64 years old) who were enrolled in the PPMI study soon after

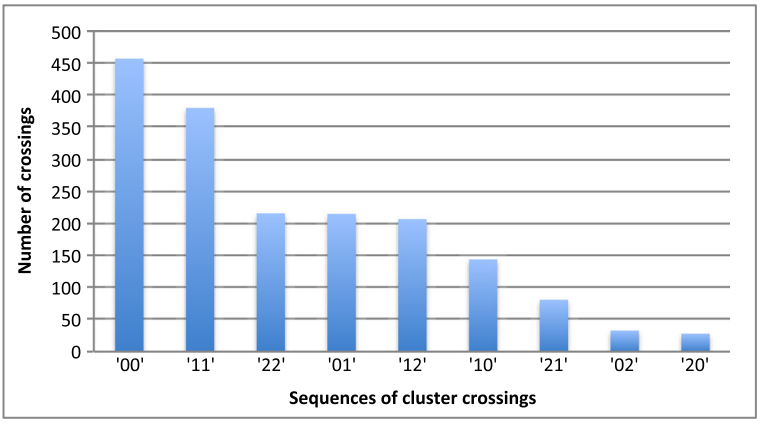


Fig. 5 Histogram resulting from 3-skip-4-gram analysis. The possible cluster crossings are listed on the X-axis and the Y-axis represents the number of cluster crossings

their Parkinson's disease diagnosis (in less than 6 months). A common thread of these patients is that they have had problems with anxiety at some point of the disease (quantified with score 1 – a symptom is present and has a slight influence on the patient's quality of life). Most of these patients have also started feeling a decline in their cognitive functions. These patients were treated with the combinations of dopamine agonists and MAO-B inhibitors. When patients motor symptoms have slightly worsened, the clinicians have tried to stabilize them by increasing the dosage of dopamine agonists, changing the dopamine agonists medication, or in rare cases introducing levodopa. These treatments are in accordance with the new practices for Parkinson's disease—clinicians introduce MAO-B inhibitors to protect the neural system of the patient, and prescribe dopamine agonists in order to control motor symptoms that are bothering the patients, and in that way they prolong the time before levodopa is introduced in the therapy.

An inspection of patients with cluster change pattern 2000 reveals that two patients who exhibit this pattern are elder female patients (more than 71 years old), with two years between the time of their diagnosis and their enrolment into the PPMI study. In the time of their initial visits, both patients had problems with their facial expression, problems with fingerntapping, hand movement, pronation-supination, toe-tapping, leg agility, postural tremor, rest tremor amplitude and constancy of rest tremor. For both patients, these symptoms were prominent on the left-hand side. In addition to their motor problems, both patients have experienced problems with depression and anxiety. Patients' medications log reveals that once the patients' motor symptoms were deemed problematic (at that time point the patients were assigned to *Cluster 2*), their respective clinicians started the symptoms treatment with levodopa medications. The introduction of levodopa lead to stabilization of the symptoms, and in our research, we observe a crossing of the patients from *Cluster 2* to *Cluster 0*.

6 Conclusions

The aim of our research is to develop a methodology which will make suggestions to the clinicians about the possible treatment changes that will improve the patient's quality of life. We also aim to discover groups of patients that follow interesting patterns of symptoms change in hope that their disease progression will reveal common symptoms and medications threads, which could benefit the future patients. Our methodology contains tracking the changes in medication patterns, clustering, rule learning and skip-grams. The results confirm known facts about the Parkinson's disease: the motor symptoms, tremor, shaking, involuntary movement, etc. are the characteristic symptoms of the disease and significantly affect the quality of life of the suffering patient. We show that Parkinson's disease patients can be divided into clusters ordered in accordance with the severity of their symptoms. By following the evolution of symptoms for each patient separately, we were able to determine patterns of medications change which can lead to the improvement or worsening of the patients' quality of life.

We introduced skip-grams as a method for following the progression of the disease. The analysis showed that the progression of the disease is mostly steady in the period of five years involvement in the PPMI study—the patients stay in the initially assigned clusters or they move to the adjacent clusters. Analysis of 3-skip-4-grams outlined groups of patients

with interesting patterns of cluster changes. We detected a group of older patients, who were not treated for a longer period and whose treatment consists of direct introduction of levodopa for treatment of motor symptoms. The other interesting group are younger patients, who were recently diagnosed with Parkinson's disease and whose treatment included the combinations of MAO-B inhibitors and dopamine agonists. In further work, we will consult medical experts for specific patients with interesting sequences and ask them to interpret their etymological characteristics, motor symptoms, and changes of therapy.

Results from the multi-view clustering setting are underwhelming in terms of the quality of produced clusters. However, the results reveal the importance of autonomic symptoms to the quality of life of Parkinson's disease patients.

The rules describing the obtained clusters were either very general (merged view setting) or very specific (multi-view setting) and may not be of sufficient assistance to clinicians. This is due to the nature of the used data, i.e. a vector of attribute sums (merged view) or a high-dimensional vector of attributes with numeric values. In future work, we will test our methodology with only a handful of carefully chosen attributes. These attributes, selected with the help of Parkinson's disease specialists, will be described by nominal values used in the clinicians' everyday practice i.e. normal, non-problematic, problematic. We believe that by an expert-assisted decrease of feature space dimensionality, we will be able to obtain descriptions of groups of patients which are even more meaningful and helpful to the clinicians. Additionally, we will improve the medications suggestion process to produce numerical suggestions of medications dosages which should be prescribed to the patients. An interesting direction for further work is to explore other clustering approaches, in particularly hierarchical clustering. Attributes from the MDS-UPDRS and MoCA questionnaires can be ordered hierarchically and exploiting this characteristic may lead to better defined groups of patients with similar symptoms. Transitions between such clusters could reveal more specific and detailed patterns of disease progression. Besides skip-grams we plan to explore other possibilities to handle temporal data. For example, we want to compare the state of a patient in a given time point with all of its past time points (not only the immediately preceding one).

Acknowledgements This work was supported by the PD_manager and HBP SGA1 projects, funded within the EU Framework Program for Research and Innovation Horizon 2020 grants 643706 and 720270, respectively. We acknowledge also the support of the Slovenian Research Agency (research core funding P2-0103 and P2-0209).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners. Corporate Funding Partners: AbbVie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, GLAXOSMITHKLINE (GSK), Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery (MSD), Pfizer Inc, Piramal Imaging, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB. Philanthropic Funding Partners: Golub Capital. List of funding partners can be also found at www.ppmi-info.org/fundingpartners.

Appendix A: Comparison of clustering algorithms on merged data set

We considered three clustering approaches for the merged data set: k-means, k-medoids, and DBSCAN. We clustered the merged data into different number of clusters and evaluated the quality of the produced clusters with the internal cluster validity metrics: SA (Rousseeuw

Table 5 Cluster validation measures for k-means, k-medoids, and DBSCAN, where k presents the number of clusters. Clustering was performed on the merged data set. Better clusters quality is marked with higher values of SA and CH, and lower values of DB

	k-means			k-medoids			DBSCAN		
k	SA	DB	CH	SA	DB	CH	SA	DB	CH
2	0.516	0.916	162.540	0.505	0.918	162.539	-0.362	0.996	16.946
3	0.368	0.916	162.540	0.336	0.918	162.539	-0.132	0.996	16.946
4	0.371	1.263	54.103	0.318	1.387	54.099	0.250	0.796	297.712
5	0.287	1.151	40.546	0.259	1.235	40.546	nan	nan	inf
6	0.275	1.256	32.412	0.253	1.283	32.412	nan	nan	inf
7	0.284	1.619	26.991	0.253	1.364	26.990	nan	nan	inf

1987), DB (Davies and Bouldin 1979), and CH (Caliński and Harabasz 1974). Table 5 presents the results of cluster validation for the selected clustering methods and the chosen number of clusters. The results show that the best performing approach is k-means.

Appendix B: Features selected by unsupervised feature selection

We used unsupervised feature subset selection to select the most relevant attributes for clustering algorithms. We used the SPEC algorithm (Zhao and Liu 2007) implemented in

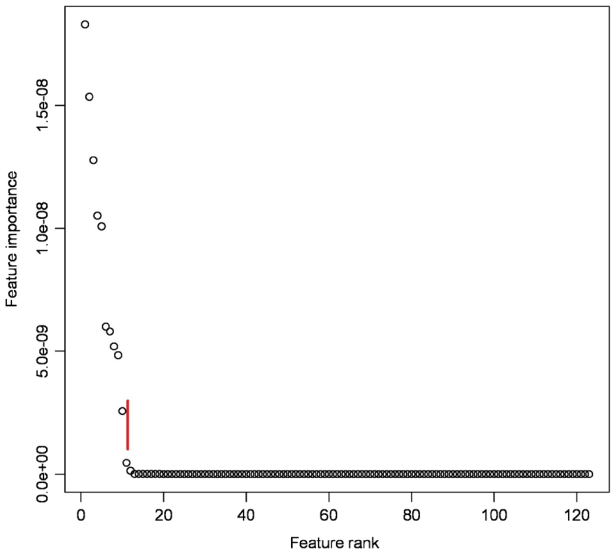


Fig. 6 Attribute rank vs attribute importance as determined by the SPEC algorithm (the most influential attribute has rank 1)

Table 6 The most important attributes ordered according to SPEC (see Fig. 6)

Attribute	Attribute description	Data set
MCAREC4	Delayed recall - daisy	MoCA
NHY	Hoehn and Yahr score	MDS-UPDRS Part III
NP3PTRML	Postural tremor (left hand)	MDS-UPDRS Part III
NP3SPCH	Speech problems	MDS-UPDRS Part III
NP2EAT	Eating tasks	MDS-UPDRS Part II
NP1SLPD	Daytime sleepiness	MDS-UPDRS Part Ip
NP3RIGLU	Rigidity (left arm)	MDS-UPDRS Part III
NP1PAIN	Pain and other sensations	MDS-UPDRS Part Ip
NP3FTAPL	Finger tapping (left hand)	MDS-UPDRS Part III
NP3RTCON	Constancy of rest	MDS-UPDRS Part III

Python (Li et al. 2016). Figure 6 presents the evaluation of attributes relevance. Based on the results, we selected the attributes left from the red line in Fig. 6. This resulted in a list of 10 attributes, presented in detail in Table 6.

In Table 7 we present the cluster validation values on the data set containing only the best attributes (listed in Table 6). The results reveal that the merged data set (consisting of sums of attributes) produces better quality clusters than the data set reduced with feature subset selection.

Results from Tables 5 and 7 show that better clusters are produced when sums of attribute values from the considered views are used as attributes in the merged data set. Parkinson's disease patients experience a whole range of symptoms, both motor and non-motor, and it is tougher for traditional clustering algorithms to separate them into groups of similar patients. The introduction of sums makes it possible to have a view of the overall status of the patients concerning particular sets of symptoms (i.e. motor symptoms, non-motor symptoms, autonomic symtptoms etc.).

Table 7 Cluster validation measures for k-means, k-medoids, and DBSCAN, where k presents the number of clusters. Clustering was performed on the data set containing only attributes from Table 6

	k-means			k-medoids			DBSCAN		
k	SA	DB	CH	SA	DB	CH	SA	DB	CH
2	0.379	1.199	102.657	0.379	1.199	102.657	0.379	1.199	102.657
3	0.337	1.199	102.657	0.283	1.199	102.657	nan	1.199	102.657
4	0.296	1.590	34.168	0.217	1.781	34.168	nan	nan	inf
5	0.279	1.580	25.608	0.170	1.745	25.607	nan	nan	inf
6	0.267	1.617	20.471	0.189	1.649	20.471	nan	nan	inf
7	0.262	1.694	17.046	0.182	1.969	17.047	nan	nan	inf

Appendix C: Evaluation of multi-view clusterings

In order to determine how the choice of data sets influence the results of multi-view clustering, we executed multi-view clustering on all 21 pairs of views, i.e. $\frac{7 \cdot 6}{2}$ pairs. Clusters resulting from each pair were evaluated using SA (Rousseeuw 1987) and the results are presented in Table 8. SA is a normalized value (range from -1 to 1) and is used to compare cluster quality on these data sets. Since clustering was performed on different data sets (each pair is effectively a different data set) and values of DB and CH are not comparable across data sets, we do not present these values. The value of each cell in Table 8 corresponds to the quality of clusters obtained by multi-view clustering on the data sets from the corresponding row and column. For example, SA (Rousseeuw 1987) on clusters obtained by multi-view clustering on the MDS-UPDRS Part I (NUPDRS1) and MoCA is 0.021. The best cluster is marked with bold.

The results show that all pairs produce clusters with low quality, but the three best performing pairs according to SA are: (SCOPA-AUT, MDS-UPDRS Part II), (MDS-UPDRS Part III, MDS-UPDRS Part II), and (PASE, MDS-UPDRS Part II).

We used the Adjusted Rand Index (ARI) (Hubert and Arabie 1985) to compare cluster structures discovered by different cluster configurations. The value of ARI is 0 for two random clusterings and 1 for two identical clusterings. Table 9 presents the ARI score computed on pairs of the winning two-view clustering settings. Results reveal that all pairs of clusterings are quite similar, and the (NUPDRS3, NUPDRS2P) and (PASE, NUPDRS2P) pairs produce almost identical clusters ($ARI = 0.966$). As the quality of individual pairs is rather low (see Table 8), there is little chance that further combinations of views would improve the quality.

Nevertheless, we constructed two additional settings for multi-view clustering by systematically adding views (data sets) to the winning bi-view clustering setting (SCOPA-AUT, MDS-UPDRS Part III). We in turn added the remaining data sets from the second (MDS-UPDRS Part II and MDS-UPDRS Part III) and third (PASE and MDS-UPDRS Part III) best performing bi-view clustering setting, thus obtaining two new multi-view settings: (SCOPA-AUT, MDS-UPDRS Part II, MDS-UPDRS Part III) and (SCOPA-AUT, MDS-UPDRS Part II, MDS-UPDRS Part III, PASE). We evaluated the quality of clusters produced by these three settings and presented the results in Table 10, where we also

Table 8 Value of SA on clusters discovered with multi-view clustering on pairs of data sets. Higher values of SA indicate clusters with better quality

	MOCA	NUPDRS1	NUPDRS1P	NUPDRS2P	NUPDRS3	PASE
NUPDRS1	0.021					
NUPDRS1P	0.023	0.014				
NUPDRS2P	0.022	0.033	0.024			
NUPDRS3	0.025	0.038	0.015	0.168		
PASE	0.023	0.059	0.013	0.162	0.048	
SCOPA-AUT	0.024	0.018	0.013	**0.173**	0.047	0.031

Table 9 ARI scores for the best performing pairs of two-view multi-view clusterings

	(NUPDRS3, NUPDRS2P)	(PASE, NUPDRS2P)
(SCOPA, NUPDRS2P)	0.488	0.504
(NUPDRS3, NUPDRS2P)		0.966

included the cluster quality measures when all views are considered and the scores of the best single view clustering on the merged data set. Please note that since clustering was performed on different data sets, values of DB and CH are not comparable. SA is a normalized value (range from −1 to 1) and is used to compare cluster quality on these data sets.

Based on the SA values from Table 10, clustering with the best clustering is produced on the merged data set that consists only of sums of attribute values from 7 data sets from Section 3.3. In the multi-view setting, best results were obtained when three data sets were considered (SCOPA-AUT, MDS-UPDRAS Part II, MDS-UPDRS Part III). The SCOPA-AUT data set contains attributes describing the autonomic symptoms of patients. The MDS-UPDRS Part II data expresses 'motor experiences of daily living', including speech problems, the need for assistance with the daily routines such as eating or dressing, etc, while the MDS-UPDRS Part III data set describes the motor symptoms which are the most characteristic symptoms of Parkinson's disease. Even though the clusters produced by the multi-view setting are of lower quality than those produced on the merged data set, results from Table 10 reveal that it might be beneficial to combine multiple data sets: the inclusion of the MDS-UPDRS Part III data set in the best performing bi-view clustering setting (SCOPA-AUT, MDS-UPDRS Part II) (SA = 0.173) produces clusters with an improved quality (SA = 0.205). These results also show that the inclusion of other, seemingly uncorrelated data sets (PASE, MOCA, MDS-UPDRS Part I, MDS-UPDRS Part Ip) can lead toward significant decrease in the quality of clusters.

In addition to the work presented above, we also used unsupervised feature subset selection to select the most relevant attributes from each of the seven views (data sets). We evaluated the quality of clusters on the newly generated data sets following the procedure presented in this section. Results showed that the quality of the clusters in these new set-

Table 10 Comparison of cluster quality using silhouette analysis (SA) for different setting of multi-view clustering

The *Data set* column presents the symptoms data sets that are used in the multi-view clustering

Data set	SA
SCOPA, NUPDRS2P	0.173
SCOPA, NUPDRS2P, NUPDRS3	0.205
SCOPA, NUPDRS2P, NUPDRS3, PASE	0.0195
All 7 data sets	0.0514
Merged data set	0.347

tings was significantly lower than the quality of clusters presented here. For that reason we did not include this part of research into the paper.

Appendix D : Rules describing multi-view clusters

We present rules describing clusters obtained by multi-view clustering using three views (SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III) i.e. the best multi-view clustering according to SA from Table 10. Attributes with the prefix SCAU are symptoms from the SCOPA-AUT data set. The suffix in the names of these attributes designates the nature of the autonomic symptoms. Attributes SCAU1-SCAU7 describe gastrointestinal symptoms, urinary problems are recorded by attributes SCAU8-SCAU13, while attributes SCAU14-SCAU16 hold information about patient's cardiovascular problems. Attributes SCAU17-SCAU18, SCAU20-SCAU21 describe thermoregulatory problems, while attribute SCAU19 describes any pupillomotor issues that a patient might be experiencing. Attribute prefixes determine the data set of their origin. Attributes with prefix NP2 are from the MDS-UPDRS Part II, while the prefix NP3 designates attributes from the MDS-UPDRS Part III data set (including attributes NHY and DYSKPRES).

Tables 11, 12, and 13 present rules describing *cluster 0*, *cluster 1*, and *cluster 2* respectively, obtained by multi-view clustering. Rules are induced on the data set that is a concatenation of the three views: SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III. Contrary to the rules obtained by the single view clustering on the merged data set where groups of patients were described by the severity of their overall status, the multi-view clusters are described by symptoms. These rules mostly describe the motor status of Parkinson's disease patients (attributes from MDS-UPDRS Part III), and are supported by their motor ability in daily living (attributes from MDS-UPDRS Part II) and their autonomic symptoms (SCOPA-AUT).

Table 11 Description rules for *cluster 0* of the multi-view clustering approach generating clusters with best quality. Views were represented by the SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III data sets

Rule		p	n
IF:			
SCAU19 ≤ 2 AND NP3PSTBL ≤ 2 AND NP3HMOVL > 2 AND			
NP3GAIT ≤ 1 AND NP3RTARU ≤ 0	$\leftarrow$ cluster $= 0$	58	0
ELSE IF:			
NP3RISNG ≤ 1 AND NP3FTAPR ≤ 0 AND NP3HMOVL > 0 AND			
NP3HMOVR ≤ 0 AND NP3RTARU ≤ 0 AND NP3FACXP ≤ 2	$\leftarrow$ cluster $= 0$	146	8
ELSE IF:			
NP3FTAPL > 1 AND NP2FREZ ≤ 0 AND NP3RTARU ≤ 0 AND			
NP3RTALU > 0	$\leftarrow$ cluster $= 0$	87	0
ELSE IF:			
NP2SALV ≤ 3 AND NP3FRZGT ≤ 0 AND NP3FTAPL > 2 AND			
NP3LGAGL > 0	$\leftarrow$ cluster $= 0$	25	4

Table 12 Description rules for *cluster 1* of the multi-view clustering approach generating clusters with best quality. Views were represented by the SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III data sets

Rule		p	n
IF:			
NP3RIGLU ≤ 0 AND NP3RIGN ≤ 1 AND NP3RTARU > 1	$\leftarrow$ cluster $= 1$	143	3
ELSE IF:			
NP3RTCON > 2 AND SCAU18 ≤ 1 AND NP2SUM ≤ 15 AND NP3RTARU > 1 AND NP3FACXP ≤ 2	$\leftarrow$ cluster $= 1$	83	3
ELSE IF:			
NP3RIGLL ≤ 0 AND NHY ≤ 1 AND NP3RTARU > 0 AND SCAU6 ≤ 1	$\leftarrow$ cluster $= 1$	40	6
ELSE IF:			
NP3PRSPL ≤ 0 AND NP3RTCON > 1 AND NP3RTARU > 0	$\leftarrow$ cluster $= 1$	37	2
ELSE IF:			
SCAU12 ≤ 1 AND NP3RTALL ≤ 0 AND NP2TRMR > 0 AND NP2EAT ≤ 0 AND NP3HMOVL ≤ 0 AND SCAU20 ≤ 0 AND NP3RTARU > 0 AND SCAU7 ≤ 0	$\leftarrow$ cluster $= 1$	15	3
ELSE IF:			
NP3RIGLU $=$ (0,1] AND NP3RTCON > 1 AND SCAU17 ≤ 1 AND NHY ≤ 2 AND NP3RTARU > 1	$\leftarrow$ cluster $= 1$	19	4
ELSE IF:			
NP3RTCON > 2 AND NP2HWRT > 0 AND NP3LGAGL ≤ 0 AND NP3TTAPL $=$ (0,1] AND SCAU6 ≤ 1	$\leftarrow$ cluster $= 1$	7	2
ELSE IF:			
NP2SALV ≤ 0 AND NP3RIGLU ≤ 0 AND SCAU17 ≤ 1 AND NP2WALK ≤ 0 AND NP3RTARL > 0	$\leftarrow$ cluster $= 1$	5	2
ELSE IF:			
NP2EAT > 0 AND NP3GAIT ≤ 0 AND NP3RTARU > 0 AND NP2SPCH ≤ 0 AND SCAU4 ≤ 1	$\leftarrow$ cluster $= 1$	6	1
ELSE IF:			
SCAU18 > 0 AND NP2DRES > 0 AND NP3SPCH ≤ 0 AND NP3RTARU > 1 AND NP3RTALU ≤ 0	$\leftarrow$ cluster $= 1$	4	0

Table 13 Description rules for *cluster 2* of the multi-view clustering approach generating clusters with best quality. Views were represented by the SCOPA-AUT, MDS-UPDRS Part II, and MDS-UPDRS Part III data sets

Rule		*p*	*n*
IF:			
SCAUSUM > 5 AND NP3RTCON ≤ 0 AND NP3FTAPR > 2	← cluster = 2	36	1
ELSE IF:			
NP3RTCON ≤ 0 AND NP2TRMR ≤ 0 AND NP3TTAPL ≤ 0	← cluster = 2	53	1
ELSE IF:			
NP3RTCON ≤ 0 AND NP3HMOVR > 0 AND NP3TTAPL ≤ 0	← cluster = 2	52	0
ELSE IF:			
NP3RTCON ≤ 0 AND NP3PTRML ≤ 0 AND NP3TTAPR > 1	← cluster = 2	73	12
ELSE IF:			
NP3RTCON = (0,1] AND NP3HMOVR > 2 AND NP3TTAPR > 0	← cluster = 2	15	1
ELSE IF:			
NP3RTALJ ≤ 0 AND NP3GAIT ≤ 1 AND SCAU8 > 2 AND NP3RTALU ≤ 0 AND SCAU4 ≤ 1	← cluster = 2	13	2
ELSE IF:			
NP3RTCON ≤ 0 AND NP2WALK > 0 AND NP3PSTBL ≤ 2 AND NP2HWRT > 0 AND NP2SWAL > 1 AND SCAU6 ≤ 2	← cluster = 2	10	0
ELSE IF:			
NHY > 1 AND NP3KTRML ≤ 0 AND PN3RIGRL > 1 AND NP3RTARU = (0,1]	← cluster = 2	38	23
ELSE IF:			
NP3RTCON ≤ 0 AND NP3HMOVL ≤ 0 AND NP3RTARU ≤ 0 AND NP3TTAPR > 0	← cluster = 2	23	0
ELSE IF:			
NP3PSTBL > 2 AND NP2DRES > 0 AND NP3RTARU > 1	← cluster = 2	7	0
ELSE IF:			
NP3RTCON = (0,1] AND NHY > 1 AND NP3FTAPR > 1 AND NP3RTARL ≤ 0 AND NP2HOBB ≤ 1 AND NP2SPCH > 0 AND NP3FACXP > 0 AND NP3RTALU ≤ 0	← cluster = 2	24	11
ELSE IF:			
NP3PRSPL ≤ 0 AND NP3RTCON ≤ 0 AND NP3POSTR ≤ 2 AND SCAU18 ≤ 1 AND NP3RIGLL ≤ 0 AND NP3RTARU ≤ 0 AND SCAU3 ≤ 0	← cluster = 2	20	4

References

Agrawal, R., Imieliński, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In *ACM sigmod record* (Vol. 22, pp. 207–216). ACM.

Appice, A. (2017). Towards mining the organizational structure of a dynamic event scenario. *Journal of Intelligent Information Systems*, 1–29.

Appice, A., & Malerba, D. (2016). A co-training strategy for multiple view clustering in process mining. *IEEE Trans Services Computing*, *9*(6), 832–845.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256.

Bence, J.R. (1995). Analysis of short time series: correcting for autocorrelation. *Ecology*, *76*(2), 628–639.

Bezdek, J.C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on computational learning theory, ACM, New York, NY, USA, COLT'* (Vol. 98, pp. 92–100).

Broder, A.Z., Glassman, S.C., Manasse, M.S., Zweig, G. (1997). Syntactic clustering of the web. *Computer Networks and ISDN Systems*, *29*(8-13), 1157–1166.

Cai, X., Nie, F., Huang, H. (2013). Multi-view k-means clustering on big data. In *Proceedings of the 23rd international joint conference on artificial intelligence IJCAI 2013, Beijing, China, August 3-9, 2013* (pp. 2598–2604).

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1–27.

Ceravolo, R., Rossi, C., Kiferle, L., Bonuccelli, U. (2010). Nonmotor symptoms in Parkinson's disease: the dark side of the moon. *Future Neurology*, *5*(6), 851–871.

Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K. (2009). Multi-view clustering via canonical correlation analysis. In *Proceedings of the 26th annual international conference on machine learning, ICML 2009* (pp. 129–136).

Choi, E., Schuetz, A., Stewart, W.F., Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, *24*(2), 361–370.

Cleuziou, G., Exbrayat, M., Martin, L., Sublemontier, J. (2009). CoFKM: a centralized method for multiple-view clustering. In *Proceedings of the ninth IEEE international conference on data mining (ICDM 2009), miami, florida, USA, 6-9 December 2009* (pp. 752–757).

Dalrymple-Alford, J., MacAskill, M., Nakas, C., Livingston, L., Graham, C., Crucian, G., Melzer, T., Kirwan, J., Keenan, R., Wells, S., et al. (2010). The moCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology*, *75*(19), 1717–1725.

Davies, D.L., & Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227.

De Alba, E., Mendoza, M., et al. (2007). Bayesian forecasting methods for short time series. *The International Journal of Applied Forecasting*, *8*, 41–44.

De Vine, L., Zuccon, G., Koopman, B., Sitbon, L., Bruza, P. (2014). Medical semantic similarity with a neural language model. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management* (pp. 1819–1822). ACM.

Dorsey, E., Constantinescu, R., Thompson, J., Biglan, K., Holloway, R., Kieburtz, K., Marshall, F., Ravina, B., Schifitto, G., Siderowf, A., et al. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, *68*(5), 384–386.

Ernst, J., & Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, *7*(1), 191.

Ernst, J., Nau, G.J., Bar-Joseph, Z. (2005). Clustering short time series gene expression data. *Bioinformatics*, *21*(suppl 1), i159–i168.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD* (Vol. 96, pp. 226–231).

European Parkinson's Disease Association (2016). http://www.epda.eu.com/, accessed: 2016/07/01.

Foltynie, T., Brayne, C., Barker, R.A. (2002). The heterogeneity of idiopathic Parkinson's disease. *Journal of Neurology*, *249*(2), 138–145.

Gamberger, D., & Lavrač, N. (2002). Expert-guided subgroup discovery: methodology and application. *Journal of Artificial Intelligence Research*, *17*, 501–527.

Gatsios, D., Rigas, G., Miljkovic, D., Seljak, B.K., Bohanec, M. (2016). m-health platform for Parkinson's disease management. In *Proceedings of 18th international conference on biomedicine and health informatics CBHI*.

Gil, D., & Johnson, M. (2009). Diagnosing Parkinson by using artificial neural networks and support vector machines. *Global Journal of Computer Science and Technology*, *9*(4), 63–71.

Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., et al. (2008). Movement Disorder Society-sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disorders*, *23*(15), 2129–2170.

Goetz, C.G., Luo, S., Wang, L., Tilley, B.C., LaPelle, N.R., Stebbins, G.T. (2015). Handling missing values in the MDS-UPDRS. *Movement Disorders*, *30*(12), 1632–1638.

Greene, D., Doyle, D., Cunningham, P. (2010). Tracking the evolution of communities in dynamic social networks. In *Advances in social networks analysis and mining (ASONAM), 2010* (pp. 176–183). IEEE.

Guthrie, D., Allison, B., Liu, W., Guthrie, L., Wilks, Y. (2006). A closer look at skip-gram modelling. In *Proceedings of the 5th international conference on language resources and evaluation (LREC-2006)* (pp. 1–4).

He, X., Kan, M.Y., Xie, P., Chen, X. (2014). Comment-based multi-view clustering of Web 2.0 items. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 771–782): ACM.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology Neurosurgery & Psychiatry*, *55*(3), 181–184.

Imhoff, M., Bauer, M., Gather, U., Löhlein, D. (1998). *Time series analysis in intensive care medicine*. Tech. rep. SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

Kaufman, L., & Rousseeuw, P. (1987). *Clustering by means of medoids*. North-Holland.

Kaufman, L., & Rousseeuw, P.J. (1990). *Finding groups in data. an introduction to cluster analysis*. Wiley Series in Probability and Mathematical Statistics Applied Probability and Statistics.

Kumar, A., & III, H.D. (2011). A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th international conference on machine learning, ICML* (pp. 393–400).

Lewis, S., Foltynie, T., Blackwell, A., Robbins, T., Owen, A., Barker, R. (2005). Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(3), 343–348.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H. (2016). Feature selection: a data perspective. arXiv:160107996.

Lin, J., Keogh, E., Wei, L., Lonardi, S. (2007). Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, *15*(2), 107–144.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J. (2010). Understanding of internal clustering validation measures. In *Proceedings of IEEE 10th international conference on data mining (ICDM)* (pp. 911–916).

Liu, Y., Li, W., Tan, C., Liu, X., Wang, X., Gui, Y., Qin, L., Deng, F., Hu, C., Chen, L. (2014). Meta-analysis comparing deep brain stimulation of the globus pallidus and subthalamic nucleus to treat advanced Parkinson disease: a review. *Journal of Neurosurgery*, *121*(3), 709–718.

Ma, L.Y., Chan, P., Gu, Z.Q., Li, F.F., Feng, T. (2015). Heterogeneity among patients with Parkinson's disease: cluster analysis and genetic association. *Journal of the Neurological Sciences*, *351*(1), 41–45.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al. (2011). The Parkinson's Progression Markers Initiative (PPMI). *Progress in Neurobiology*, *95*(4), 629–635.

Michalski, R.S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence*, *20*(2), 111–161.

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:13013781.

Minarro-Giménez, J.A., Marín-alonso, O., Samwald, M. (2013). Exploring the application of deep learning techniques on medical text corpora. *Studies in Health Technology and Informatics*, *205*, 584–588.

Murugesan, S., Bouchard, K., Chang, E., Dougherty, M., Hamann, B., Weber, G.H. (2017). Multi-scale visual analysis of time-varying electrocorticography data via clustering of brain regions. *BMC Bioinformatics*, *18*(6), 236.

National Collaborating Centre for Chronic Conditions. (2006). *Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care*. London: Royal College of Physicians.

Patel, S., Lorincz, K., Hughes, R., Huggins, N., Growdon, J., Standaert, D., Akay, M., Dy, J., Welsh, M., Bonato, P. (2009). Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, *13*(6), 864–873.

PD_manager: m-Health platform for Parkinson's disease management (2015). EU Framework Programme for Research and Innovation Horizon 2020, Grant number 643706, 2015–2017. http://www.parkinson-manager.eu/.

Ramani, R.G., & Sivagami, G. (2011). Parkinson disease classification using data mining algorithms. *International Journal of Computer Applications*, *32*(9), 17–22.

Reijnders, J., Ehrt, U., Lousberg, R., Aarsland, D., Leentjens, A. (2009). The association between motor subtypes and psychopathology in Parkinson's disease. *Parkinsonism & Related Disorders*, *15*(5), 379–382.

Riviere, C.N., Reich, S.G., Thakor, N.V. (1997). Adaptive Fourier modeling for quantification of tremor. *Journal of Neuroscience Methods*, *74*(1), 77–87.

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Samà, A., Pérez-López, C., Rodríguez-martín, D., Moreno-aróstegui, J.M., Rovira, J., Ahlrichs, C., Castro, R., Graċa, R., Guimarães, V., et al. (2015). A double closed loop to enhance the quality of life of Parkinson's disease patients: REMPARK system. *Innovation in Medicine and Healthcare*, *2014*(207), 115.

Schieb, L.J., Mobley, L.R., George, M., Casper, M. (2013). Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and healthcare characteristics. *Stroke*, *44*(1), 146–152.

SENSE-PARK (2016). Project's website: http://www.sense-park.eu/, accessed: 2016/07/01.

Stecher, J., Janssen, F., Fürnkranz, J. (2014). Separating rule refinement and rule selection heuristics in inductive rule learning. In *Proceedings of machine learning and knowledge discovery in databases - European conference, ECML PKDD 2014* (pp. 114–129).

Szymański, A., Kubis, A., Przybyszewski, A.W. (2015). Data mining and neural network simulations can help to improve deep brain stimulation effects in Parkinson's disease. *Computer Science*, *16*(2), 199.

Timmer, J., Gantert, C., Deuschl, G., Honerkamp, J. (1993). Characteristics of hand tremor time series. *Biological Cybernetics*, *70*(1), 75–80.

Tzallas, A.T., Tsipouras, M.G., Rigas, G., Tsalikakis, D.G., Karvounis, E.C., Chondrogiorgi, M., Psomadellis, F., Cancela, J., Pastorino, M., Waldmeyer, M.T.A., et al. (2014). PERFORM: a system for monitoring, assessment and management of patients with Parkinson's disease. *Sensors*, *14*(11), 21,329–21,357.

Tzortzis, G., & Likas, A. (2009). Convex mixture models for multi-view clustering. In *Proceedings of the 19th international conference artificial neural networks - ICANN* (Vol. 2009, pp. 205–214).

Valmarska, A., Robnik-Šikonja, M., Lavrač, N. (2015). Inverted heuristics in subgroup discovery. In *Proceedings of the 18th international multiconference information society*.

Valmarska, A., Miljkovic, D., Lavrač, N., Robnik-Šikonja, M. (2016). Towards multi-view approach to Parkinson's disease quality of life data analysis. In *Proceedings of the 5th international workshop on new frontiers in mining complex patterns at ECML-PKDD2016*.

Valmarska, A., Lavrač, N., Fürnkranz, J., Robnik-Šikonja, M. (2017). Refinement and selection heuristics in subgroup discovery and classification rule learning. *Expert Systems with Applications*, *81*, 147–162.

Visser, M., Marinus, J., Stiggelbout, A.M., Van Hilten, J.J. (2004). Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. *Movement Disorders*, *19*(11), 1306–1312.

Washburn, R.A., Smith, K.W., Jette, A.M., Janney, C.A. (1993). The physical activity scale for the elderly (PASE): development and evaluation. *Journal of Clinical Epidemiology*, *46*(2), 153–162.

Xu, C., Tao, D., Xu, C. (2013). A survey on multi-view learning. *Neural Computing and Applications*, *23*(7–8), 2031–2038.

Zhao, J., Papapetrou, P., Asker, L., Boström, H. (2017). Learning from heterogeneous temporal data in electronic health records. *Journal of Biomedical Informatics*, *65*, 105–119.

Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th international conference on machine learning* (pp. 1151–1157). ACM.