Title: The Prediction of Leadership Degree Based on Machine Learning

Author: Olena Skrynnyk, Tetyana Vasilyeva

1. Introduction

Digital mapping of issues lacking directly measurable characteristics is a scientific challenge, regardless of the nature of the process. In order to analyse complex social, organisational, managerial or psychological coherences computer-aided, scientists resort to different approaches: indirect parameterisation via measurable variables, physicalisation of facts, measurement / analysis via correlations. Approaches based on artificial intelligence enabled an entirely new field of treatment.

In order to identify and eliminate the undesirable actions, incorrect decisions, inadequate motivational measures, ineffective distribution of tasks must be promptly recognised by a leader and appropriate measures must be taken. We hypothesise that for automatic decisions on whether to select a more appropriate service delivery strategy or situational appropriate participation component, the use of an AI-assisted approach, would be highly reliable. The aim of our study consisted in the search for such a solution.
The literature review revealed a research gap in this area, prompting further research to find such a solution. Thus, the purpose of our study was shifted to identifying the possibility of predicting performance as a function of leadership level using selected inputs. Our study consisted of successive steps. In the first step, we have formulated a conceptual model that utilises the measurable leadership parameters for coherence through delineation of the real leadership system (restrictions, assumptions) and formulation of the organisational model of leadership (parameters, relationships). The second step included prediction of organisational model parameters and simulation with predicted parameters. These steps and the results obtained form the focus of the following sections.

2. Literature review

The first step in finding a suitable artificial intelligence-based method for predicting the change of leadership parameters over time consisted in a thorough analysis of the publications on the topic under investigation. According to the successive literature analysis of the publications in WoS and Scopus databases, the latest developments in the last five years (2016-2020) reveal the following research directions: Cobbe, et al. and Ranjan & Foropon considered general approaches to decision-making processes supported by artificial intelligence, which offer the frameworks for the application of artificial intelligence algorithms [4], [18]; Rakhimboev et al. investigated machine learning-supported decision-making for defined organisational units as well as the whole organisation [17]. Analysis of scientific findings, resulting from the proposed issues, generated the important impulses for the procedure and formulation of the parameter model. Scientific approaches with the greatest relevance to the issue we have considered relate to the findings on decision-making of specific organisational and human resources issues. This was the main point in research of Grobelny et al., Fallucchi, Hilb, Jain et al., Puranam [7], [9], [10], [11], [19]. These publications piqued our interest from a technology deployment perspective. Furthermore, Elhag & Almarashi devoted special focus to the tangible methods and their implementations in decision-making, such as the observation of regression tree analysis application [6].

The search for analysis and prediction of leadership style, leadership culture and employee motivation with the help of artificial intelligence generated much fewer results. Bisi et al., Merrick, explored leadership and decision-making in game approaches [2], [15]. Ahmad & Akhtaruzamman investigated the impact of various leadership styles on employees [1]. Leyer & Schneider, Martin researched in the direction of applying of artificial intelligence for strategic decision-making as a supporting or stand-alone function [13], [14], [20]. Gadiraju et al. explored the further field of research extends to the AI-supported analysis of the participation patterns in the groups/organisations [8].

Analysis of listed publications points to the research gap in predicting leadership degree, despite a variety of research interests highlighting different facets of the topic “Leadership”.

3. Research methodology

Our research interest consisted in identifying extant scientific knowledge and resulting approaches for the prediction of specific leadership parameters, otherwise develop a solution to predicting these parameters. Our study relied on the methods of applied research for the delineation of the investigation space and applied existing python libraries for prediction and simulation.

We applied the method of scientific abstraction to represent real-life leadership behaviour and decision-making refined without side effects and insignificant details and have determined that a considerable part of the organisational activities regarding leadership and independent of the management level are related to parameters described as employees, company, decision and working performance. In addition, we formulated the model of the leadership degree. We have decided on leadership parameters in our model in order to differentiate staff management from pure resource management. Similarly, we have divided all parameters into organisational and characteristic model parameters. The results of this step were captured in a surrogate model and have already been published [21].

Machine learning is a sub-field of artificial intelligence. By applying machine learning, information technology systems acquire the ability to recognise patterns and regularities from existing data. The
insights from the data can be generalised and utilised for innovative problem-solving or for the analysis of previously unknown data [3].

For the analysis of disorderly scattered data, the method of regression decision tree analysis is particularly profitable. Trained regression decision trees map information as decision rules in the form of a (fictitious) more or less branched tree. These decision rules do not specify values for the independent variable, but to determine the estimated value, the algorithm follows the branching direction in the tree. Compared to linear models, the model evaluation in regression decision tree analysis is simple, but not as precise [12]. Nevertheless, the prediction is sufficient for the specified purpose.

In order to represent the correlation between the requested data sets for prediction, the training of the machine learning model is required. The precondition for achieving these goals is the subdivision of the data into at least one training data set and one model test data set. The test data set is used to analyse the performance of the machine learning algorithm. For the large data sets, the division into 80% test data and 20% training data is conventional. For reduced prediction data, the KFold cross validation method is applied [16]. This method aims to make several runs with the same data set, but to split it differently each time into a training set and a validation set. This division is performed either according to KFold Cross Validation (division of the data set into k equal-sized folds) or Leave One Out-Cross Validation (e.g. with 100 data the division is 99:1). Due to the available dataset, we applied the KFold Cross Validation method [5].

4. Machine learning approach and its parameters

There are two major blocks, offline training and online prediction to predict leadership degree. Based on organisational parameters, whose parameters are determined by data-based method, leadership degree organisational model should be developed. To achieve the approximation of the real organisational unit behaviour depth is required, as the degree of leadership varies depending on parameters (Effectiveness of Decision-Making, Effectiveness of Employees' Motivation, Effectiveness of Employees' Work, Degree of Delegation, Degree of Participation, Intensity of Leadership) in reality.

The resulting surrogate organisational model has the same organisational characteristics as the real company data, given reasonable predicted parameters. With the company data and parameters (Current performance, Rights, Degree of regulation of orderliness of work, Employee loyalty) a machine learning algorithm is trained. This allows the other parameters (Input, Number of leadership units, Span of control, Delegation rate, Participation rate, Delegation limit, Participation limit, Target performance, Target duration, Current duration) to be predicted for the unknown organisational scenarios. Then, the degree of leadership can be determined by organisational model, resulting in the simulated performance.

Organisational leadership model describes the rationality of the strategy selected and executed by the management unit in terms of resource planning and deployment, as well as staff retention and motivation measures. In this context, organisational parameters include: Current performance, Right (Rights/Sum of right), Employee loyalty and Degree of regulation of orderliness of work. These are crucial for the prediction of the characteristic model parameters (the remaining parameters). Furthermore, we have extended the organisational surrogate model by some conditions, which serves as a basis for machine learning prediction model. The generated predictive values of Effectiveness of Decision-Making and Effectiveness of Employees’ Motivation as well as Degree of Participation form the input for the calculation of Simulated Performance.

An illustration of the procedure for implementing machine learning in our grey model appears in Figure 1. The details can be retrieved from the pseudo code in section Appendix.
The organisational model relies on the correlations of constants (for a longer period of time determined parameters), parameters that change over time, and those calculated from direct causation. Table 1. Output parameters form the vector of leadership degrees and function as markers for certain leadership qualities.

<table>
<thead>
<tr>
<th>Leadership parameter</th>
<th>Input parameters</th>
<th>Output parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target performance (Pt)</td>
<td>Current performance (Pc)</td>
<td>Degree of Leadership (DL)</td>
</tr>
<tr>
<td>Input (I)</td>
<td>Current duration (dc)</td>
<td>Degree of Delegation (DD)</td>
</tr>
<tr>
<td>Target duration (dt)</td>
<td>Delegation rate (RaD)</td>
<td>Degree of Participation (DP)</td>
</tr>
<tr>
<td>Span of control (SpC)</td>
<td>Participation rate (RaP)</td>
<td>Effectiveness of Decision-Making (EDM)</td>
</tr>
</tbody>
</table>

Figure 1: System overview of the offline training and online prediction with the Decision tree model
### Input parameters

<table>
<thead>
<tr>
<th>Number leadership units (NLU)</th>
<th>Delegation limit (LiD)</th>
<th>Participation limit (LiP)</th>
<th>Working climate (WC)</th>
<th>Target performance (Pt)</th>
</tr>
</thead>
</table>

### Output parameters

<table>
<thead>
<tr>
<th>Employee loyalty ($f_e$)</th>
<th>Sum of rights ($\text{SumR}$)</th>
<th>Rights</th>
<th>Degree of regulation of orderliness of work ($f_w$)</th>
<th>Current performance ($P_c$)</th>
<th>Degree of Leadership ($DL$)</th>
</tr>
</thead>
</table>

Success of a planned activity of managed group is possible only when the absolute motivation as well as the tactical effectiveness are entirely provided. The initial evaluation of work and motivation effectiveness is based on the difference of the target/ current performance state, where the combination of parameters is determining. This difference is also crucial for the effectiveness of decision-making. However, the difference does not affect other parameters.

The test and training data originate from a retail company that previously possessed a low degree of digitalisation and managed its data in tabular and analogue form.

We used data collected from five years and eight months, excluding holiday-related short weeks. The smear was taken on the last day of the working week. Over the time period in which the data was collected, the organisation operated on a two- or three-level basis. Management (strategic and tactical) or subordinate management units (operational) made decisions. This investigation is grounded in sales data and tactical sales and personnel strategies. Input as a parameter in our example refers to the calculated costs of labor, procurement costs of units to be sold, rent, electricity and tax. Target performance as a parameter is monetized.

### 5. Results

Based on the organisational leadership model, we have conducted simulation describing different scenario alternatives with parameters Current Performance, Rights, Employee loyalty and Degree of regulation of orderliness of work and consequently data sets predicting other parameters. The simulation algorithm executed seven splits, each with 300 learning data and 42 test data. Since only the data of one company are used in our machine learning model, we set the minimum/maximum values for the parameters undefined.

For the following example, we decided to select the 16-th prediction parameter set. The example illustrates the new market entry of a product variant and predicts the parameter values with values given at the time of zero. Similar to the zero series in production, the first pieces of the new product variant are finally tested by customers before initial sales (customer interest, predictable sales quantities, pricing). Due to relatively limited sales volumes, comparably high procurement costs and the associated risk, this strategy has proven successful for small trading companies.
Figure 2: Prediction results of leadership parameters

Figure 2 compares the prediction of the parameter data with the target data (original data). The visible difference between some prediction curves and the original curves is the result of limited availability of training and test data, and thus the variance generated. The parameters Effectiveness of Employees’ Work and Effectiveness of Motivation remain comparatively the same (the change of employee motivation effectiveness is in 0.001 range and seems to grow when scaling representation). The predicted parameter values Participation and Delegation Degrees curves indicate their entry only around the eighth day of sales due to similar courses in past years. In contrast, Effectiveness of Decision-Making increases rapidly, but not to its peak level yet. The value of leadership intensity remains constant.

Although simulative performance seems to have the zero value at the beginning, due to the scale of display grid this value is at two and increases slightly (Figure 3).
Figure 3: Prediction results of leadership parameters

The selected example shows the predicted parameters of the Degree of Leadership in a defined period of time. In addition, according to the given conditions, the display of the parameter value as strings appears. To highlight the assignment of individual value levels to the subsequent measures, we specifically labelled them for this review:

\[
DL = \begin{cases} 
    EW = \text{borderline} = \text{can be improved} \\
    EM = \text{bad} = \text{should be improved} \\
    DP = \text{good} \\
    DD = \text{good} \\
    EDM = \text{bad} = \text{should be improved} \\
    InL = \text{remains constant}
\end{cases}
\]

These evaluations are necessary for the next step, the recommendation of measures.

For the purpose of statistically evaluating the obtained prediction results and thus the model in the context of used machine learning algorithm, we investigated 3-sigma areas and histogram (Figure 4). The larger and more coherent the training and test data set, the smaller the 3-sigma range. 3-sigma analysis at Effectiveness of Work and Effectiveness of Motivation indicates comparably ideal range, at Degree of Participation and Degree of Delegation the deviation is about 0.18, whereas at Effectiveness of Decision Making this range remains relatively large. For the statistical analysis of prediction of the parameter Intensity of Leadership is convenient the exploration of normal distribution difference between original and predicted for test data since compared to the other parameters Intensity of Leadership is invariant over time in a data set. The largest error value is around zero level, which indicates relatively small deviation.
Figure 4: Prediction results of leadership parameters

6. Conclusion and discussion

With the purpose of digitally mapping an organisation as a complex social living entity of human connections with certain characteristics and weighted influences, we encountered a question of direct correlation between leadership parameters and performance. We have extended the question to the possibility of prediction. In search of proper solution or a scientific approach for prediction and mapping of leadership parameters, we have encountered knowledge gap. Both scientists and internships have considered and investigated leadership from different perspectives, the holistic solution which would be applicable for companies should be designed and developed. In our investigation, we have designed a machine learning algorithm for predicting leadership degree parameters and related simulated performance.
Despite good prediction and relatively minor errors, our algorithm has some limitations: first, the low number of learning and test data, which causes that the algorithm lacks diversity of possible developments.

On the other hand, since the data sets of a small company were chosen, the applicability of the algorithm is not possible without additional simulations with data from medium and large companies, as well as with other sales volumes.

The improvement potentials lie in the reliability of prediction (also for the larger datasets with one of the additional methods: Gaussian Process, Gradient Boosting, k-Nearest Neighbors, Multi-layer Perceptron, Multi Target), extension of the parameter model by complex processes and decisions (by approaching comprehensive hybrid stepwise approaches of machine learning).

We envision the extensive application of machine learning-based models for human-related parameters as part of future management systems, which will use predictive mechanisms to analyse and target the activities and behaviour of employees and managers.

Furthermore, the results of our study present great scientific interest due to the potentially enhancing the organisational theoretical perspective of leadership by leadership degree, its measurement and prediction approaches.

7. Acknowledgements

At this point we would like to express our special thanks to company Delait for trust and sharing their organisational and performance data for testing and training of machine learning model. We would also like to thank Oleksandr Marchuk for his competent efforts in programming some code passages.

The survey was supported by the Ministry of Education and Science of Ukraine and performed the results of the project “Modeling and forecasting of the socio-economic-political road map of reforms in Ukraine for the transition to a sustainable growth model” (registration number 0118U003569).

8. References


9. Appendix

# begin of pseudo code: ---------------------------------------
class Predict_Orga ():
    def __init__(self, ML_model, debug is False, normalize is True):
        def normalize_x (self, x):
            return x
        def fit (self, x, y):
            self.ML_model.fit (x, y)
        def predict (self, x):
            return Predict_OUTPUT, param_OUTPUT
        def InL (self, Nlu, SpC):
            return InL
        def DD (self, RaD, t, LiD, SpC, right):
            return DD
        def DP (self, RaP, t, LiP, SpC, right):
            return DD
        def EM (self, Pc, t, Pt, f_e, f_p):
            return EM
        def EW (self, DP_EW, t, Pt, LiP, SpC, right):
            return EW

# end of pseudo code: ---------------------------------------
```python
    return DP_EW
def EDM(self, d_c, t, d_t, Pc, I, Pt):
    return EDM
def DL(self, DP, EW, model_parameter, orga_parameter):
    return DL
def Ps(self, d_c, d_t, EDM, DP, I, EW):
    return Ps

# simulate function
    def simulate_DD (self, param, orga):
        return DD
    def simulate_DP (self, param, orga):
        return DP
    def simulate_EM (self, param, orga):
        return EM
    def simulate_EW (self, param, orga):
        return EW
    def simulate_EDM (self, param, orga):
        return EDM

orga_model is Predict_Orga (debug isTrue)
dir is os.getcwd ()
model_parameter is pd.read_csv (os.path.join (dir, 'Orga_Parameter.csv'))

# pass the crash parameter
OrgaParameter are model_parameter # Pc, right, f_e, f_p

# pass the model parameter
ModelParameter are model_parameter # Nlu, SpC, RaD, LiD, RaP, LiP, Pt, d_c, d_t, I

InL_error is empty list for InL

DD_error is empty list for DD error
DD_origin is empty list for DD origin data

DP_error is empty list for DP error
DP_origin is empty list for DP origin data

EM_error is empty list for EM error
EM_origin is empty list for EM origin data

EW_error is empty list for EW error
EW_origin is empty list for EW origin data

EDM_error is empty list for EDM error
EDM_origin is empty list for EDM origin data

# permutation of data
Permutation is permutation(OrgaParameter)

# KFold cross validation
kf is KFold (n_splits is 7)
kf.get_n_splits (Permutation)
```
KFold (n_splits is 7, random_state is None, shuffle is False)

for train_index, test_index in kf.split (Permutation):
    x_train, y_train are OrgaParameter (train_index), ModelParameter (train_index)
    x_test, y_test are OrgaParameter (test_index), ModelParameter (test_index)

    # training from the model
    orga_model.fit (x_train, y_train)

    for i, orga, param in zip(range(x_test), x_test, y_test):
        # prediction of velocity
        Predict_OUTPUT, param_OUTPUT is orga_model.predict (orga)

        # calculate of degree of leadership
        DP is orga_model.simulate_DP (param, orga)
        EW is orga_model.simulate_EW (param, orga)
        DL is orga_model.DL (DP, EW, param, orga)

        # call of InL
        InL_ORIGIN is orga_model.InL (param, param)
        InL_PREDICT is Predict_OUTPUT from 0

        # call of DD
        DD_ORIGIN is orga_model.simulate_DD (param, orga)
        DD_PREDICT is Predict_OUTPUT from 1
        DD_origin.append (DD_ORIGIN)

        # call of DP
        DP_ORIGIN is orga_model.simulate_DP (param, orga)
        DP_PREDICT is Predict_OUTPUT from 2
        DP_origin.append (DP_ORIGIN)

        # call of EM
        EM_ORIGIN is orga_model.simulate_EM (param, orga)
        EM_PREDICT is Predict_OUTPUT from 3
        EM_origin.append (EM_ORIGIN)

        # call of EW
        EW_ORIGIN is orga_model.simulate_EW (param, orga)
        EW_PREDICT is Predict_OUTPUT from 4
        EW_origin.append (EW_ORIGIN)

        # call of EDM
        EDM_ORIGIN is orga_model.simulate_EDM (param, orga)
        EDM_PREDICT is Predict_OUTPUT from 5
        EDM_origin.append (EDM_ORIGIN)

        # calculate of simulated performance
        Ps_PREDICT is empty list
        for k, edm, dp, ew in zip (range(EDM), EDM, DP, EW):
            d_c is param_OUTPUT from 7
            d_t is param_OUTPUT from 8
            I is param_OUTPUT from 9
            Ps is orga_model.Ps (d_c, d_t, edm, dp, I, EW)
            Ps_PREDICT.append (Ps)

    # end of pseudo code: ------------------------------------------