Machine learning as a tool for choice of enterprise development strategy

M Krichevsky^{1*}, and *E Serova*²

¹State University of Aerospace Instrumentation, 67, Bolshaya Morskaya st., St. Petersburg, 190000, Russia

²National Research University Higher School of Economics, 3, Kantemirovskaya st., St. Petersburg, 194100, Russia

Abstract. One of the main objectives of strategic management is the development and selection of strategies to achieve the desired results. The main goal of this paper is the analysis of the main domains or areas of machine learning application to support the process of strategic planning and decision making. The scientific methodology of the research studies is methods and procedures of modeling and intelligent analysis. This is theoretical and empirical paper in equal measure. This paper deals with the issues of machine learning implementation and how intellectual models and systems can be used to support the process of strategic planning. At the preprocessing stage on the basis of a modeled base of examples of strategy options, the use of clustering methods for forming groups of similar parameters that influence the choice of strategies and groups of similar enterprise objects, each of which has a certain type of strategy, is demonstrated. On the next step the selection of ranked characteristics that affect the choice of strategy is made. At the stage of solving the problem of choosing strategies, module Classifier Learner from MatLab 2018b is used.

1 Introduction

This paper represents an extended and largely revised report of the authors [1]. For any company, the strategy selection is an important task, which determines further activities. In fact, the strategy is a long-term particular direction of company's development in terms of the internal relationship system, the company's position among competitors. In other words, the strategy is a set of decision-making rules applied by the company in its activities [2]. Each of the strategies selected by the company leads to different consequences. That's why the strategies shall be selected based on all the available information at the decision-making stage. According to [2], the strategy is an elusive and abstract concept; but it is the tool that can significantly help the company in unstable conditions. As a result, the strategy as a management tool serves as a critical reference point for the companies involved in various areas of activity.

One of the essential decision-making tools in various areas of activity is machine learning, which is becoming an increasingly popular concept in the modern world since its most

* Corresponding author: mkrichovsky@mail.ru

common goal is to optimise systems by allowing one to make smarter use of products and services [3]. In the manufacturing industry machine learning can lead to cost savings, time savings, increased quality and waste reduction. At the same time, it enables systems to be designed for managing human behaviour.

A complexity of business dynamics often forces decision-makers to make decisions based on subjective mental models, reflecting their experience. However, research has shown that companies perform better when they apply data-driven decision-making. This creates an incentive to introduce intelligent, data-based decision models, which are comprehensive and support the interactive evaluation of decision options necessary for the business environment [4].

In recent years, Machine Learning models have been replacing the traditional methods of data processing in management. For example, intensive studies of bankruptcy prediction models for credit risk management have shown that the bankruptcy assessment methods based on machine learning models (support vector machines, bagging, boosting, and random forest) give more accurate forecast evaluation than those previously adopted. [5].

The article is further organised as follows. First, machine learning methods that can be applied to solve the strategy selection problem are considered. Then data pre-processing using cluster analysis is briefly described, the method of dimensionality reduction is specified, tools used in the implementation of company development strategies are analysed, assessments for the selection of the company management method using the Classifier Learner module are given. The final part of the article contains the discussion of the results obtained, and the directions of further research.

2 Methodology

The analysis and handling of any data imply model building based on the observations and its further use, for example, in classification, forecasting, etc. The methods used in artificial intelligence (AI) for data handling are included in machine learning, which is a subset of AI. The term “machine learning” means the technology defined in [6] as follows: “Optimisation of the model performance criterion using data and past experience”. We have the model defined with the accuracy of some parameters, and learning is the execution of the computer program for optimisation of model parameters. The model can be predictive (to make predictions in the future) or descriptive (to gain knowledge from data), or both. In machine learning (ML) data plays an essential role, and the learning algorithm is used to discover and get knowledge or properties from data. The quality or quantity of the dataset will affect the learning and forecasting efficiency.

Machine learning uses the theory of statistics when building mathematical models, because the primary task is inferencing based on data sampling. Computer science plays a dual role. First, learning requires the efficient algorithms to solve optimisation problems, as well as to store and process the huge amount of available data. Second, after the model study, its presentation and the algorithmic decision for inference shall also be efficient. In some applications, the efficiency of the learning or inference algorithm, i.e., its spatial and time complexity, can be as important as the prediction accuracy.

The name of ML reflects the fact that the described method analyses data and independently finds the model without human intervention. This process is called learning because it resembles learning with data to find a model. Therefore, the data used in ML is

learning data. ML methods are currently used in various fields, for example, in management for risk assessment, in production, development of recommendation systems [7-9].

2.1 Cluster analysis

Cluster analysis belongs to uncontrolled (unsupervised) methods and represents an easy way to identify homogeneous groups of objects called clusters. Objects (or observations) of a particular cluster have similar characteristics, but differ significantly from the objects not belonging to this group. Here the Self-Organising Map (SOM) is used for clustering. Its main purpose is to divide samples into groups (clusters) by certain features. The notable characteristic of this algorithm is that the input vectors that are close — similar — in high dimensional space are also mapped to nearby nodes in the 2D space. It is in essence a method for dimensionality reduction, as it maps high-dimension inputs to a low (typically two) dimensional discretised representation and conserves the underlying structure of its input space [10].

SOM differs from typical artificial neural network both in its architecture and algorithmic properties. Its structure comprises of a single-layer linear 2D grid of neurons, instead of a series of layers. All the nodes on this grid are connected directly to the input vector, but not to one another, meaning the nodes do not know the values of their neighbours, and only update the weight of their connections as a function of the given inputs. The grid itself is the map that organises itself at each iteration as a function of the input of the input data. As such, after clustering, each node has its own (i,j) coordinate, which allows one to calculate the Euclidean distance between 2 nodes.

The process of learning of the network of this type is as follows. The vector from the learning sample is provided to the input. For this vector, the winning cluster is determined and the weight coefficients are adjusted so that to make it closer to the input vector. Most often, the weights of several neighbours of the winning cluster are also adjusted. In contrast to supervisory learning, the purpose of this type of learning is not to compare the network inference with the required one, but to adjust the weights of all connections to ensure the maximum coincidence with the input data.

SOMs have many advantages, the most important of which probably is the ease of interpretation, and they have seen many different applications in very diverse areas. Even huge data sets can be tackled: training may be slow in these cases, but once this is done, new data can be mapped to the SOM very quickly. However, they do have some disadvantages, as well. The first and foremost is the number of parameters that has to be set. The size and topology of the map needs to be determined, as well as the values for the training parameters. Furthermore, since the process has several random components (the initialisation, the order of the presentation of the data), maps with identical topology and training settings may show quite different results. In many cases, however, the conclusions drawn from these seemingly different maps may still be similar.

2.2 Feature selection

One of the common tasks of ML is the selection of predictors from a large list of candidates. The large number of input parameters makes the learning of neural networks difficult and increases time expenditure. The term curse of dimensionality generally refers to the difficulties involved in fitting models, estimating parameters, or optimising a function in

many dimensions. As the dimensionality of the input data space (i.e., the number of predictors) increases, it becomes exponentially more difficult to find global optima for the parameter space, i.e., to fit models. Hence, it is simply a practical necessity to screen and select from among a large set of input (predictor) variables those that are of likely utility for predicting the outputs (dependent variables) of interest.

The methods implemented in the Feature Selection and Variable Screening module of Statistica 13 program are specifically designed to handle extremely large sets of continuous and/or categorical predictors, for regression or classification-type problems. You can select a subset of predictors from a large list of candidate predictors without assuming the relationships between the predictors and the dependent variables. Therefore, this module can serve as an ideal in ML, to select manageable sets of predictors for further analysis.

For continuous predictors, the program will divide the range of values in each predictor into k intervals (10 intervals by default; to “fine-tune” the sensitivity of the algorithm to different types of monotone and/or non-monotone relationships, this value can be changed by the user in the Feature Selection and Variable Screening Startup Panel). The continuous variables are recoded with a special algorithm. Categorical predictors will not be transformed in any way.

2.3 Classification tools

2.3.1 Decision trees.

Decision trees are used to predict the placement of classes or objects in the category based on the results of parameters measured [11-12]. Typically, decision-making is the implementation of methods to perform binary classification by splitting tree branches based on the measured values of the predictors. The leaves of such trees are class labels, and the branches are conjunctions of features that lead to these class labels. The end node contains the answer. Upon classification, trees give nominal answers, such as true or false.

Classification trees are widely used in diagnosing medical cases, data structure formation, and decision making. The graphical format of results provides better interpretation. The purpose of formation of decision trees is to predict the values of the dependent variable as a function of the corresponding values of one or more independent variables (predictors).

2.3.2 Linear discriminant analysis.

Discriminant analysis (DA) is a classification method. DA models have been developed from the middle of the last century. Significant progress in their development is associated with the names of foreign scientists R. Fischer and H. Hotelling. Despite the long history of such methods, DA models are still being actively used in various fields of activity [13-14].

DA implementation assumes that the data comes from the Gaussian mixture model. In this case, DA is expected to be a good classifier. Besides, this method assumes by default that the covariance matrices of all classes are equal.

Linear DA is a generalisation of Fisher's Linear Discriminant, i.e. the method used in statistics, pattern recognition, and machine learning to find a linear combination of features that characterises or separates two or more classes. For learning (creating) the classifier, the parameters of the Gaussian distribution of each class are assessed. When predicting the

classes of new data, the trained classifier finds the class with the lowest misclassification value.

2.3.3 *Support vector machines*

Support Vector Machines and the related kernel methods perform classification by constructing an N-dimensional hyperplane that optimally separates data into two categories. Support Vector Machine (SVM) models are closely related to neural networks. In fact, the SVM model based on the sigmoid kernel function is equivalent to the perceptron consisting of two layers of neurons.

This method was developed by the Russian scientists V.N. Vapnik and A.Y. Chervonenkis, who at the end of the 70s of the last century developed the theory of the generalised portrait, which describes the method of finding the separating hyperplane in the linear classification problem [15,16]. Their theory is often called VC-theory, especially in foreign publications. It characterises the properties of the machines trained, allowing them to properly generalise new data.

In the simplest classification the SVM uses the linear separating hyperplane to create the classifier with a maximal margin. In this case, the learning problem is reduced to the constrained non-linear optimisation problem.

If the data cannot be separated in the original feature space, the SVM first performs non-linear transformation of the original feature space into a higher dimensional space. This operation can be performed using various non-linear transformations: polynomials, multilayer perceptrons, etc. After that, the SVM's task related to finding the linear optimal separating boundary in the new feature space becomes trivial.

2.3.4 *Nearest neighbour method*

The k-nearest neighbours (k-NN) algorithm is a controlled machine learning algorithm used to solve classification problems. This method assumes that similar objects are located next to each other. The k-NN algorithm reflects the idea of similarity, sometimes called distance or proximity, and stipulates the calculation of the distance between points on the graph [14,17].

In this method, input data consists of the k nearest learning examples in the feature space. Output data depends on whether the k-NN algorithm is used for classification or regression. Upon classification, output data represents the class members. The object is classified by a set of votes of its neighbours, and the object is assigned to the class most common among its k nearest neighbours (k is a positive, usually small integer). If $k = 1$, the object is simply assigned to the class of that single nearest neighbour.

At the classification stage, the k parameter is a user-defined constant, and the unlabelled vector (query point or breakpoint) is classified by assigning the label that is most common among the k learning samples closest to that query point. Euclidean distance is the distance metric that is typically used for continuous variables.

For the new instance, class membership predictions are made by searching the most similar instances (neighbours) in the entire k learning set, and summing the output variable for these k instances.

3 Results

Various company development strategies were presented back in 1957 by I. Ansoff. However, until now, this strategic management tool has been actively used to determine the main directions of business development [18]. The most critical business and development strategies of the company include:

- Penetration strategy.
- Market development strategy.
- Product development strategy.
- Diversification strategy.

Each of the strategies is characterised by the following indicators: x1 - expansion of the market share; x2 - increase in the number of purchases of goods; x3 - increase in the frequency of purchases of goods; x4 – discovering new opportunities for product use; x5 - use of new distribution channels; x6 – searching and winning new market segments; x7 – product sales in new regions; x8 – upgrading of existing products; x9 - expansion of the product range; x10 - creation of the new generation (models) of products; x11 - development and manufacture of a fundamentally new product; x12 - presence of competence for the development of new business; x13 – growth opportunity in the current markets with the current staff.

These indicators determine the tools that influence the strategy selection. The set of these tools may differ slightly in different sources [18, 19], but this factor is not of fundamental importance for this paper.

To create the example database for neural network training, two approaches can be used:

- Use of real data.
- Use of toy datasets.

In machine learning, it is important to learn how to use toy datasets correctly, since algorithm learning using real data is difficult and may fail [19]. Toy datasets are critical to understand the algorithm work. A simple synthetic data sample makes it quite easy to assess whether the algorithm has learned the necessary rule or not. This assessment using real data is difficult. Here the Monte Carlo method is used to generate toy data sample.

Based on the analysis of the proposed strategies and features that affect these strategies, we assume as follows:

- To implement strategy 1 (penetration), high values of features x1 - x4 are required.
- To implement strategy 2 (market development), high values of features x5 - 8 are required.
- To implement strategy 3 (product development), high values of features x1, x9 - x11 are required.
- To implement strategy 4 (diversification), high values of features x4, x12, x13 are required.

Each feature was generated according to the uniform distribution law, based on the 10-point scale, taking into account high and low values of the parameters. There was the total of 40 options of implementation, so the observation matrix consists of 40 rows and 13 columns. In addition, random noise was added to the draw values. The obtained values of 40 objects with different strategies are shown in Figure 1 in the three-dimensional space of the principal components.

As can be seen from Figure 1, objects do not explicitly form 4 clusters, although the degree of data loss upon transfer from 13 features to 3 aggregated features is about 17%, which is an acceptable value, given the sample size.

Fig. 1. Modelled objects in the principal component space.

Table 1 shows the calculated values of parameters upon transfer to the principal components.

Table 1. Parameters of the principal components.

Eigenvalue	% Total variance	Cumulative %
2. 806	46.67	46.77
1.177	19.62	66.39
1.007	16.79	83.18

The next step of the analysis includes the selection of features using the Feature Selection and Variable Screening module, which allows to rank features in order of importance. Figure 2 shows the graph that allows to select the most critical features.

Fig. 2. Histogram of the most important predictors.

The final selection of features is the analyst's task. For the current task, we take 6 features numbered 13, 7, 12, 2, 11, 10 for further research.

Then we'll carry out cluster analysis using the Kohonen network for all objects with such 6 features. The Self Organising Maps (SOM) method is one of the options of clustering of multidimensional vectors – the projection algorithm that preserves topological similarity. When using this algorithm, vectors that are close in the original space become close on the resulting map.

The learning map can be represented as a puff cake, each layer of which is coloured based on one of the features of the initial data. The resulting set of colourings can be used to analyse the patterns between the dataset components. After map formation, a set of nodes which can be displayed as a two-dimensional image is obtained. The analysis result is shown in Figure 3.

Fig. 3. Formed clusters of objects.

As can be seen from Figure 3, two clusters numbered 1 and 4 form compact groups, the remaining two clusters numbered 2 and 3 are not so compact. Note that during the use of this method in this problem, the number of clusters was previously assumed to be 4.

Now let's look at the classification of such objects (recall that each object is a company with a certain type of strategy) performed using the Classification Learner module from the MatLab 2018b program. This module can work in the manual mode, when one method or another is checked in turn, and in the automated mode, when the program studies all the possible solutions for this problem and indicates the best results. We note again that machine learning methods based on the available data form the model that can be used to classify new observations. Here we use the first (manual) mode, which is more visual, according to the authors.

Figure 4 shows the results of use of four classification methods: decision trees, discriminant analysis, support vectors, and k-nearest neighbours.

▼ History		
1 ☆ Tree	Accuracy: 75.0%	
Last change: Disabled PCA		6/6 features
2 ☆ Linear Discriminant	Accuracy: 92.5%	
Last change: Linear Discriminant		6/6 features
3 ☆ SVM	Accuracy: 95.0%	
Last change: Linear SVM		6/6 features
4 ☆ KNN	Accuracy: 90.0%	
Last change: Fine KNN		6/6 features

Fig. 4. Classification results.

As can be seen from Figure 4, Support Vector Machines methods demonstrates the best accuracy among the four specified methods. However, it should be noted that the accuracy estimate depends on various conditions, for example, the selected predictors (their composition can be easily changed), the nature of the error matrix (the matrix indicates the areas of poor work of the classifier), the area under the receiver operating characteristic (the area proximity to 1 indicates an error-free classification).

Below are the results for the method selected by the program. Figure 5 shows the graph of the analysis objects in the plane of the first two parameters: Varname 1 (growth opportunity in the current markets with the current staff) and Varname 2 (product sales in new regions).

Fig. 5. Prediction results.

After completion of classifier training based on the available sampling, the graph displays the prediction results of the selected model (Support Vector Machine). Crosses in the figure 5 indicate erroneous results.

Figure 6 shows the confusion matrix for this method. The confusion matrix rows are the true classes; the columns are the predicted classes. The diagonal elements show the coincidence of the true and predicted classes. As can be seen from the matrix, the trained classifier correctly sorted only two classes of objects: the second and the third ones.

The confusion matrix has the added True Positive Rates (TPR) and False Negative Rates (FNR) on the right. The last two columns on the right contain the summary data for each class. For example, the first line contains 90% and 10% for the first class, which corresponds to TPR and FNR. In other words, for the first class the share of correctly rated objects is 90%, and the share of false objects is 10%. It can also be seen that the objects of the second and third classes are rated correctly: TPR = 100%.

Fig. 6. Confusion matrix.

Figure 7 shows a receiver operating characteristic curve that shows the percentage of TPR as a function of the proportion of FPR for the selected classifier.

Fig. 7. Receiver operating characteristic.

The marker shows the FPR and TPR values for the currently selected classifier. For example, the FPR zero value indicates that the current classifier correctly assigns all observations to the positive class (when constructing an operational characteristic, with the first class selected as positive, and remaining classes selected as negative).

To predict the class of the new observation, it is necessary to export the created model to the workspace. As a result, the trainModel structure is formed, which is used for forecasting with new data. Below is an example of the result of classification of the new vector [1.0 2.0 3.0 4.0 5.0 6.0]:

```
>> ks6 = [1.0 2.0 3.0 4.0 5.0 6.0]
ks6 = 1    2    3    4    5    6
>> yfit = trainedModel.predictFcn(ks6)
yfit = 3.
```

As can be seen, the selected classifier assigns the new object to the third category.

4 Discussion

The decision method selection results obtained in the manual mode of selection of the type of classifiers can be expanded upon transfer to the automatic selection of classifiers. In the future, it is reasonable to consider for this problem the construction of the generalised assessment of the selection of strategies or analyse the features that affect the strategies. The method used in this paper for a particular problem is obviously implemented in other management and economic problems.

5 Conclusion

Thus, the paper demonstrates the use of machine learning methods based on the MatLab 2018b software product for the selection of decisions. Machine learning is an integral element of artificial intelligence, so the approach used in the work can serve as an initial element upon transition to the digital economy.

References

1. Krichevsky M and Serova E 2019 Machine Learning Methods for Strategic Management Proc. Int. May Conf. on Strategic Management (Beograd) 15(1) 422
2. Ansoff I 2007 Strategic management: Classic edition (New York: Palgrave Macmillan)
3. Anli du Preez and Gert Adriaan Oosthuizen 2019 Procedia Manufacturing 33 810
4. MarkoBohanec, MirjanaKljajić Borštnar and MarkoRobnik-Šikonja 2017 Expert Systems with Applications 71(1) 416
5. Flavio Barboza, Herbert Kimura and Edward Altman 2017 Expert Systems with Application 83 405
6. Alpaydin E 2010 Introduction to machine learning, 2nd Edition, (Adaptive Computation and Machine Learning) (Massachusetts Institute of Technology)

7. Chandrinou S K, Sakkas G, and Lagaros N D 2018 Expert Systems with Application 105(9) 34
8. Stanula P, Ziegenbein A, and Metternich J 2018 Machine learning algorithms in production: A guideline for efficient data source selection Proc. CIRP 78 261
9. Portugal I, Alencar P, and Cowan D 2018 Expert Systems with Application 97(5) 205
10. Gabbiani Fabrizio and James Cox Steven 2007 Mathematics for Neuroscientists (Elsevier Inc. Academic Press)
11. Azad M and Moshkov M 2017 European Journal of Operational Research 263(3) 910
12. Kim K and Hong J A 2017 Pattern Recognition Letters 98 39
13. Gao W, Alsarraf J, Moayedi H et al 2017 Applied Soft Computing 84(11) 105748
14. Ezghazi S, Zahi A, and Zekoua K 2017 Expert Systems with Applications 80(1) 58
15. Vapnik V and Chervonenkis A 1974 Pattern Recognition Theory (Moscow: Nauka)
16. Vapnik V 1998 Statistical Learning Theory (NY: J. Wiley)
17. Zhang X, Li Y, and Zhang X 2017 Pattern Recognition 62 33
18. Ansoff Matrix and Development Business Strategy <https://blog.iteam.ru/matritsa-ansoffa-i-strategii-rosta-biznesa>
19. Ansoff's matrix: example, construction and strategies <http://galyautdinov.ru/post/matrica-ansoffa>
20. Ramsundar B and Zadeh R 2018 TensorFlow for Deep Learning (Beijing-Boston: O'Reilly)