Fides, Ratio et Patria. Studia Toruńskie 15/2021
ISSN 2392-1463

Rafał Lizut*
Katolicki Uniwersytet Lubelski Jana Pawła II w Lublinie
The Dnipropetrovsk State University of Internal Affairs, Dnipro
ORCID: 0000-0002-6067-1469

AI IN DETERMINING INDICATORS FOR FX PREDICTION MODELS

Streszczenie

Sztuczna Inteligencja w wyznaczaniu wskaźników modeli predykcyjnych na rynkach FX

Systemy wspomagania decyzji w handlu na rynku walutowym okazały się przydatne, jeśli nie niezbędne. Stopień ich skuteczności predykcyjnej zależy od skonstruowanego modelu i dobranych danych. Podstawowe informacje do budowy modeli pochodzą z historii kursu interesującej nas pary walutowej w określonych przedziałach czasowych. Reszta danych pochodzi z wybranych wskaźników określających, co może być przyczyną takich właśnie wahań kursowych walut w danym czasie. Wskaźniki te pochodzą z różnych źródeł, takich jak teorie ekonomiczne, doświadczenia traderów, itp. Analiza różnych metod wykazała, że chociaż autorzy i praktycy stosują różne metody w przewidywaniu kursów walutowych, w tym ML, ANN i innych, nie używają żadnej z nich do weryfikacji wartości samych wskaźników lub określenia, które powinny być brane pod uwagę jako dane wejściowe dla algorytmów.

Takie podejście do wyboru wskaźników jest niewystarczające. Najlepsze modele predykcyjne byłyby oparte na związku przyczynowo-skutkowym, a nie koincydencji. Aby uniknąć błędu *non causa pro causa* należy określić relację między wskaźnikami. Jednak na skomplikowanym rynku walutowym ważne wskaźniki mogą być na pierwszy rzut oka niewidoczne, ze względu na nasze ograniczenia komputacyjne i te związane z pamięcią. W związku z tym, zaproponowane zostały metody wyszukiwania wzorców, zastosowane z powodzeniem w innych obszarach badań jak wykrywanie oszustw i korupcji, gdzie ważne wskaźnikami do modeli predykcyjnych zostały okryte przez metodologie takie jak uczenie nienadzorowane. Ponadto zasugerowano, że wskaźniki

* Rafał Lizut – Assistant Professor (IT), John Paul II Catholic University of Lublin, Poland (Department of Applied Computer Science, Faculty of Natural Sciences and Health); Professor (Economy & Management), Dnipropetrovsk State University of Internal Affairs, Dnipro, Ukraine (Departament of Analitical Economy and Management).

te mogą być określane w czasie rzeczywistym i stać się częścią adaptacyjnego systemu predykcyjnego FX.

Słowa kluczowe: AI; Sztuczna Inteligencja; Machine Learning; Uczenie Maszynowe; FX; Przewidywanie kursów walutowych; Forex

Summary

AI in determining indicators for FX prediction models

Decision support systems in FX trade has proven successfully has proven useful if not necessary. Their rate of prediction success depends on the model constructed and data selected. The basic information for constructing models comes from the interesting pair FX rate in the specific time periods. The rest of data comes from selected indicators stating what could be the reason for particular FX rates in the time. The indicators come from various sources like the economic theories, traders' experiences, etc. Analysis of various methods revealed, that although authors and practitioners employ various methods including ML, ANN and others, they do not use any of the methods to verify or determine which indicators should be considered as an input. This approach to selection of indicators is insufficient. The best prediction models would be bases on the cause-effect relation, and not the coincidental one. To avoid a non causa pro causa and other fallacies the relation the indicators need to be determined. Yet in the complicated FX market, important indicators may be invisible at first glance, due to our computational and memory limitations. Therefore some methods of pattern search including unsupervised learning are suggested due to their successes in areas like fraud and corruption detection. They have proven, that sometimes unexpected influential indicators may be found by methodologies like unsupervised learning. Also, a suggestion, that those indicators can be determined real-time and become a part of adaptive FX predictive system is made.

Keywords: AI; Machine Learning; FX; FX indicators prediction; Forex

1. Introduction

Automatic and semiautomatic systems of decision support and executing, data processing and analysis became necessary when human abilities have reached its limits. Our limitations concern the ability to process data, but also the minimum time to execute a decision. Even the time between making a decision to press a button and actually pressing it takes a significant amount of time from the machine response time perspective. For certain activities, like high-frequency stock trading, a time shorter than a blink of an eye decides whether one earns or

loses money[1]. Not only the response time is a limitation, but also human brains works slower and makes more mistakes when tired.

There is also another limitation, crucial for this article, namely the number of factors possible to be taken into account simultaneously. For example, in stock exchange market one may consider factors like weather changes or political situation in a small country on the other side of the world. Yet only certain number of factors our intellect can take into account at once. So, which should be considered and which neglected? Moreover, there exist the so-called emergent properties, which only belong to an entity created out of parts, While the parts do not have the property of the entity. For example, lines forming of triangle do not have one third properties of the triangle-ness so creation a triangle-ness does not come from adding properties of lines creating the triangle. Triangle-ness only appears after we consider three connected lines and a new entity. Using our recent pandemic situation it do not influence our situation, including the FX one, only as a collection of sick people, but new properties like social unrest and global lockdown appeared. In order to see all important indicators that influence the situation one needs to it both as a whole with various levels of precision and also as a collection of parts. It is equally important to be able to distinguish indicators of much relevance from the rest before our simulations become too complex. It is impossible to consider all factors at once so we select the relevant ones. However, in order to objectively determine which factors are important ones and which are not would have to… determine all factors at once. So, the circle closes.

A vision of controlling and predicting the future by calculating all possible factors, today we would call it a brute force method, was tempting for or ancestors. For example, there was the Laplace, who said "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."[2] Such intellect was called a Laplace's Demon. This Demon represents what is expected from a perfect decision supporting and executing system. Not necessarily the mechanistic way as Laplace was seeing it in the spirit of his time, but the whole idea is very much close to ours.

We do not have (yet) a machine which would provide such a memory and computational power. Yet from our experience we know, that the actions are

[1] Por. Lewis M., *Flash Boys: A Wall Street Revolt*, New York – London 2014.

[2] P. S. de Laplace, *A Philosophical Essay on Probabilities*, p. 4.

performed by human beings and quite often successful. In life we use heuristics, act on the basis of not fully verbalized, or even available complete set of premises. It is difficult, especially in the decisions too complexed or too fast why a leap--of faith like decision is made. Yet we have to act sometimes in such manner otherwise our life would be very slow and uneventful, the decisions and actions would take too much time. And making no decision is also a decision and has consequences. Therefore, decisions and actions are based on the so-called tacit knowledge, background believes, made via intuition being more or less vague analogy from previous experience and so one. Looking from a profession point of view this would be called craft.

The other side is science and its approach. There is continuous debate about what scientific knowledge and method is, but I will avoid going into the debate since it is not significant for the thesis provided here. But to understand what does it mean to base a selection of indicators on, let us say, scientific economic theory at least a simple explanation must occur. What I understand by scientific knowledge is a socially and historically accumulated set of justified true beliefs (I am aware of the Goettier's and others objections therefore I have indicated I will not get into the debate since it is an economical paper)of a certain type. There is a set of conditions, like being general, that knowledge must fulfil to be numbered among scientific. For this paper it is enough to say that it indicates certain regularity and this regularity is subjected to corroboration and being rejected sometimes after a long period of time. In an experimental science a theory is being created and later tested. The theory is therefore limited by the author(s) intellectual skills.

This introduction going beyond the scope of economics and IT was to shed a light on what I believe our decision making systems are missing right now. But in order to present what they are missing we need to say what is currently available.

2. Algorithms and IT in Decision Making

Computer processing have changed the game. The limitation of the number of factors considered at the same time is put away comparing to the human ones. Computers have accuracy, processing speed of serial operations, simultaneous (parallel/multithread) operations and huge memory bank which does not degrade with time or scope. Speed, parallel processing and memory allow analyzing large amount of data almost instantaneously and in rigorous manner. So the calculations are precise, repeatable and fast. This is a huge improvement but they still need to be programmed and inserted by a human with all limitations

described above. Firstly, a standard programming is slow. It is great when one can spend a lot of time to code and then this code is used many times later but what if the situation constantly requires the code change. Secondly, all the principles are explicitly written by a programmer. So, if a programmer or a group of programmers cannot comprehend the complexity of the phenomena, they cannot also write an adequate code. This obstacle was to a certain extend diminished by introducing new methodologies like neural network – form the framework perspective – and deep learning from the method perspective. In this approach it programming not so much "telling" computer what to do step by step, but to create a self-learning adaptive system that can be fed by data of accurate format[3]. And contrary to the case of human being, the more data the better, therefore there is a need for Big Data.

3. Economy, Forex and Artificial Intelligence

Economy, usually dealing with big datasets, noticed benefits of decision supporting systems and even automatic systems like automated stock trading. Also, the successful implementation with a lot of testing is being conducted for more than a decade (Hryshko and Downs[4], Koskinen[5]). Various, described below methodologies have been employed especially for stock trading. One may say, that those who do not use any technology of this type is doomed to extinct from the market. An often-used term in the application is Artificial Intelligence (AI). The term is vague enough to be applicable to most of the areas of utilization of assisting or automating technology and being fashionable provides a nice touch to commercials, articles, grant applications and so one, therefore it has also appeared in the chapter title of the article. However, if we want to talk about specifics, a particular technology or methodology needs to be addressed. Their number is vast and exceeding the scope of the article. And the goal of the article is not to determine all the types of algorithms or their implementations, but to suggest an idea to employ one of the suggested at the end or any other to determine what

[3] N. Yerashenia, A. Bolotov, D. Chan, G. Pierantoni, *Semantic Data Pre-Processing for Machine Learning Based Bankruptcy Prediction Computational Model*, https://ieeexplore.ieee.org/document/9140238/ [access: 5 Aug 2021].

[4] A. Hryshko, T. Downs, *System for foreign exchange trading using genetic algorithms and reinforcement learning*, https://www.researchgate.net/publication/43447963_System_for_foreign_exchange_trading_using_genetic_algorithms_and_reinforcement_learning [access: 5 Aug 2021], p.764.

[5] J. Koskinen, J. Airas, T. Nummelin, T. Pekkala, J. Starczewski, *Exploring algorithms for automated fx trading – constructing a hybrid model*, https://www.sciencedirect.com/topics/computer-science/markov-decision-process [access: 5 Aug 2021].

should be consider as a factor for FX predicting model. But give so some examples we have LR, SVM/SVR[6], GBC/GBR, NN[7] and others.

4. Forex, methods and indicator selection review

Forex market goal seems to be pretty straightforward. Master exchange of pair(s) in order to make profit. However, winning a Lotto prize is also simple – buy a ticket, win the prize. But the most important are prediction abilities. In winning lotto, one may try to simulate randomizing machine patterns, or detect patterns creating a "winning system", but the input data is historical data for the previous wins/loses. In Forex market there can be many influencing factors. One can try to determine the factors or/and detect a pattern. If we are talking about historical data of pairs like EUR/USD versus time if a model is built, we can only determine a pattern. For example, if EUR goes up for 6 days it drops down in the 7th. But it does not explain why it is like that. Then it is easy to mistake a cause effect relation when after occurrence of a cause there is always an effect with coincidence where phenomena occur together with some time, but they are not connected. Or to take something for a reason which is not a reason. This error is known for a long time and is called non causa pro causa. In both cases, even if the preceding phenomenon occurs, the expected one may not.

In the previous part of the paper I have indicated, that for complex systems like the economic ones, it may be difficult not only consider, but even list all influential indicators. An interesting example is the Coronavirus situation, which has influenced other factors like industry, employment and so on giving this particular FX market scenario. So far, the indicators are adopted arbitrarily[8]. Hryshko & Downs describe it in the following way "Usually traders start with a set of concepts based upon indicators and then turn those concepts into a set of rules. The rule creating process requires a subjective choice to be made of which indicators rely on and further subjectivity is needed in order to define rules for interpretation of the indicator signals. The trader then has to program the rules to create software for technical analysis." (Hryshko, Downs, 2004 p.764) [9] And it seems like a general rule for various models for creating FX prediction tools.

[6] M. Shen, C. Lee, H. Liu, P. Chang, C. Yang, *An Effective Hybrid Approach for Forecasting Currency Exchange Rates*. Sustainability 2021, https://www.mdpi.com/2071-1050/13/5/2761 [access: 5 Aug 2021].

[7] Z. Rowland, G. Lazaroiu, I. Podhorská, *Use of Neural Networks to Accommodate Seasonal Fluctuations When Equalizing Time Series for the CZK/RMB Exchange Rate*. https://www.mdpi.com/2227-9091/9/1/1 [access: 5 Aug 2021].

[8] K. Sahu, S. Nayak, H. Behera, *Forecasting currency exchange rate time series with fireworks-algorithm-based higher order neural network with special attention to training data enrichment* https://journals.agh.edu.pl/csci/article/view/3474 [access: 5 Aug 2021].

[9] A. Hryshko, T. Downs, p. 764.

The authors do not focus on the indicators adopted. At least in the initial state, some algorithms select promising indicators but only form the predetermined set. Let us provide examples for particular methods.

4.1. Markov decision process (MDP)

Simple, yet possible to formalize like Markov decision process (MDP) describing a process where the agent makes decision on the basis of observation of the environment output reward and looking at the next state to make a decision[10]. The state of the latter state of the environment is changing on the basis of the former output and reward to generate a go signal, but the elements of the system are known to the one who makes the next decision. So the system variables are known and determined and not selected by the system itself.

4.2. Heuristics

A group of heuristics like greedy search heuristics[11]. Greedy algorithm takes steps in the problem solving by looking at the nearest set of options on the path to achieving a goal an selecting the most profitable step. In the simple version the algorithm is blind to the possible steps after the one which is the nearest. In the set of the nearest steps selects the most beneficial and the process restarts. Here also each step and the state space (all possible states) are determined. Therefore both the indicators and possibilities the state can be in must be previously arbitrarily determined and introduced by the algorithm implementer. And in general it is for optimization of what was previously determined, so even if we employ it to look for the most promising indicators the indicators have to be introduced first.

4.3. Technical Analysis

Technical Analysis in short aims at prediction the future price changes on the base of the relation between time and FX prices. It tries to determine the trends and fluctuations looking for correlations between prices and a set of indicators. The analysis can provide a list of indicators with their weights (a number representing their relevance expressed in a form of a multiplier) but only from previously selected list of possible indicators to consider. And even the list should be determined by some "smart" algorithm, to avoid impossible to solve computational complexity. The general formula for the numbers of possibilities is:

[10] M. Littman, *Markov Decision Processes*, https://www.sciencedirect.com/science/article/pii/B0080430767006148 [access: 5 Aug 2021].

[11] M. Öztürk, I. Toroslu, G. Fidan, *Heuristic based trading system on Forex data using technical indicator rules*; https://www.sciencedirect.com/science/article/abs/pii/S1568494616300369 [access: 5 Aug 2021].

$$P(N, m) * 2^{m-1},$$

where:

$P(N, m)$ – is the number of permutation s of N objects taken at m time which with.

With 20 indicators (10buy and 10 sell) with the rule length of 8 indicators the number of possible rules is $6{,}5 * 10^{11}$ possibilities[12]. Therefore, the number of indicators possible form the computational point of view to consider is limited here.

4.4. Genetic Algorithm

Genetic Algorithm (GA) or evolutionary algorithm is an example of a "smarter" algorithm with its r main purpose is to optimize, so to find a minimum or a maximum of particular values or functions. They work similarly like the natural selection in Darwinian theory where next generations of selected set of indicators provides better results. For example, better approximation to the expected FX rate. Then the situation repeats and creating next generations until we are satisfied with the result, reach the designated number of the generations or the set is checked entirely (then this approach turns into a brute force search). The algorithm employs the methods of crossover in search of local minimum/maximum; mutation to search a global minimum/maximum by helping to avoid the local minimum/maximum trap, where the local minimum/maximum is considered to be the global one and the search is stopped; replication where randomly selects sets of indicators is copied to the next generation and elitism where the set(s) giving the best result is copied to the next generation. In our case the algorithm will look for the most approximate set of indicators for determining the FX rate in the future but doing it not one by one, but grouping them into sets (generations) and combining in the search of the best prediction set. In this case there is a pursuit for the search for the indicators. Here also the pursuit is only among previously selected set of indicators.

4.5. Q-learning

Q-learning offered by C. Watkins[13]. In this approach the algorithm is based on determined the most profitable indicators to represent the states of environment (in this case situation on the FX market) comparing portfolio values from the time t-1 with the time t. If the decision to let us say sell (a short position) led to diminution of the value of the portfolio, the reinforcement signal (to continue

[12] A. Hryshko, T. Downs, *ibid.,* p. 764.

[13] C. Watkins, *Learning with delayed rewards*, https://www.researchgate.net/publication/33784417_Learning_From_Delayed_Rewards [access: 5 Aug 2021].

the action or not) is negative. But here again one has to determine the indicators determined previously by let us say GA.

4.6. Reinforcement learning

There several types of reinforcement learning (RL) including the recurrent and evolutionary.

4.6.1. Recurrent Reinforcement Learning

Recurrent Reinforcement Learning (RRL) first introduced for training neural network trading systems in 1996 and then applied to FX market with its proponents Moody and Wu[14]. The goal is to search for a maximum performance function. In FX, depending on the model it can be wealth, wealth, risk-adjusted performance ratio, like Sharpe ratio, profit etc. In terms of selection of the indicators the Moody and Saffell provide an adequate description:

An example of a single asset trading system that takes into account transactions costs and market impact has following decision function:
– points) and numbered according to the following formula (style: Equation):

$$Ft = F(\Theta; Ft - 1, It) \text{ with } It = \{zt, zt - 1, zt - 2, ...; yt, yt - 1, yt - 2, ...\} = \pi r^2,$$

where:

Θt – denotes the (learned) system parameters at time t

It – denotes the information set at time t, which includes present and past values of the price series zt and an arbitrary number of other external variables denoted yt."[15]

Here the authors also talk about arbitrary number of variables, our indicators without an indication of how they are to be determined.

4.7. Evolutionary reinforcement learning

Evolutionary reinforcement learning (ERL)is explained by Koskinen et al: *"The model is initially trained for a certain period and is then applied to a test set of a certain length. After trading period, the training window is advanced by a length of certain points and the whole procedure is repeated.*"[16]. The common approach

[14] J. Moody, W. Lizhong, *Optimization of trading systems and portfolios*, https://ieeexplore.ieee.org/document/618952 [access: 5 Aug 2021].

[15] J. Moody, M. Saffell, *Reinforcement Learning for Trading*, http://papers.nips.cc/paper/1551-reinforcement-learning-for-trading.pdf Access [access: 5 Aug 2021], p. 918.

[16] J. Koskinen, J. Airas, T. Nummelin, T. Pekkala, J. Starczewski, *Exploring algorithms for automated FX trading – constructing a hybrid model*, https://www.sciencedirect.com/topics/computer-science/markov-decision-process [access: 5 Aug 2021], p. 8.

in this type of solution is that the system can select the indicators according to the environment. It can change its structure, evaluate, but only form the set of indicators previously selected by the model creator.

4.8. Hybrid approaches

Hybrid approaches like GA-RL[17] offers a solution, where the algorithm is based on determined the most profitable indicators by GA, and then uses RA to represent the states of environment (in this case situation on the FX market) comparing portfolio values from the time t-1 with the time t and depending on the comparison of portfolio value change it gives an adequate for short, long or neutral position. Here the indicators are preselected by GA, however of the set of the indicators is selected as in the described above GA model.

4.9. Artificial Neural Networks

Artificial Neural Networks (ANN). They are to emulate biological neural networks, where there exist input, output and other layers. There is a variety of the methods of organization of the ANNs, but from our perspective the most important is the input and output layer and the purpose of the ANN. In general, given the input layer the network is train to provide an expected output layer. In FX an input layer can be a historical data of currency pairs while the output data can be an expected, next-cycle pair. In the offered ANN the input type does not vary for particular ANN in terms of the preselected set of inputs.

I have presented only examples to indicate how the input data is selected. The methods here, and other that I have analyzed have in common that the method of the initial adoption of indicators is not provided *explicite* and they do not employ tools like ML to determine them. There are more systems, technologies and algorithms aiming at maximization of profit in FX market for traders like Support Vector Machines, Hidden Markov Model and more. But all of them operate on certain selected set of indicators which can be considered, ignored, reinforced, waged. But they are all preselected and limited, especially because computation complexity of the algorithm grows often exponentially with input data. Considering just few indicators more may causes huge computational complexity growth. So, increasing their number in the same stage as the described calculation is done does not seem like the best path to take.

[17] A. Hryshko, T. Downs, *ibid.*, p. 764.

5. Indicators considered

In order to improve the set of indicators let us present some of the currently used ones. Certain are necessary like pairs versus time. Moody and Saffel (2001) state, that a trading system takes historical time series FX price changes as inputs while outputs the preferred position (long or short). But there are other groups of indicators.

An important from our point of view distinction into technical and fundamental inputs.

> *"Inputs can be divided in fundamental inputs and technical inputs. Some fundamental inputs are, for example, price index, foreign reserve, GDP, export and import volume, interest rate, etc. Technical inputs are, among others, the delayed time series data, moving average, relative strength index, etc. In addition to fundamental and technical inputs, it is also possible to use individual forecast results as inputs when ANNs are used as combined forecasting tool.)"*[18].

The successful approaches from the further part of the article seem to be similar to fundamental inputs.

Another important distinction is into univariate and multivariate indicators. "Univariate inputs use data directly from the time series that is being forecast. Thus, their forecasting potential is based on the predictive capabilities of the time series itself and is related to technical analysis. Multivariate inputs use also data that is not based on the time series itself and they are used with fundamental analysis. Multivariate inputs in the context of foreign exchange rates are based on economics and finance theory. Walczak et al. state that it is necessary to use multivariate inputs to forecast exchange rates. However, according to Wei et al., most inputs in neural networks models for exchange rate prediction are univariate".[19] Here there could be an ideal space for another method to determine multivariate inputs.

Looking at the concrete indicators we have Momentum Oscillator, which together with Relative Strength Index was the basis of machine learning model for Dunis et al.[20] etc. Even though they are determined for ML, they were not

[18] W. Huang, K. Lai, Y. Nakamori, Shouyang Wang, *Forecasting Foreign Exchange Rates with Artificial Neural Networks*, https://www.researchgate.net/publication/220385306_Forecasting_Foreign_Exchange_Rates_With_Artificial_Neural_Networks_A_Review [access: 5 Aug 2021].

[19] W. Huang, K. K. Lai, Y. Nakamori, S. Wang, *Forecasting Foreign Exchange Rates With Artificial Neural Networks: A Review*, https://www.researchgate.net/publication/220385306_Forecasting_Foreign_Exchange_Rates_With_Artificial_Neural_Networks_A_Review [access: 5 Aug 2021].

[20] C. Dunis, M. Gavridis, A. Harris, S. Leong, P. Nacaskul, *An application of genetic algorithms to high frequency trading models: a case study*; https://www.researchgate.net/publica-

determined by ML. Even in the newest research where ML and other tools are so much in fashion, and where the new types of methodologies are developed, there is uses as input a predetermined set of data type. For example, Suhartono et al. aim their method to "decompose and reconstruct the time series factors which including trend, cyclic, and seasonal into several additive components, i.e. trend, oscillation and noise."[21].

To summarize, if we do not decide what to consider as input data even the best model is not going to work properly. Today's inputs in the form of the described types of variables – univariate, multivariate, technical and fundamental variables are predetermined. One can create a large set of variables and then determine using one of the described methods or others to use them, not use them or use them with wages. They are selected from the total set of possible ones in the world on the basis of economic theory, basic association like time and rate, intuition and so one. The researchers' and practitioners' effort are mainly focused on determining newer, better models of prediction employing the given set of indicators. And the set of indicators is left to the old paradigms.

6. Promising analogies for relevant indicators determination

Considering the previous parts of the article we can conclude, that also in this case we, human beings, are not capable to see all the patterns in a situation where thousands of indicators are at play. We may overlook the crucial ones. Even though our economic theory provided us with a lot of good indicators, it was also created by human beings with the described limitations. For example the emergent properties may not be perceived. The situation becomes more complicated once there is a fluctuation of those variables. Is there a way of overcoming this obstacle? An idea of possible solution I have derived from the area of combating corruption where unsupervised learning was successfully employed in the search of unforeseen correlations and indicators allowing to predict or detect corruption. There are many options for managing data like Unsupervised Learning ML family, together with Supervised Learning, Semi-supervised Learning, Reinforcement Learning, Transduction, Learning to Learn. From the perspective of this article it is only important to indicate, that certain approaches use previously categorized (labeled), selected data to operate on them, while unsupervised learning works on data which is not entirely properly labeled and finds the correlations between the elements on its own. So they can find correlations even invisible to its creator.

tion/228421645_Different_Methods_to_Clean_Up_Ultra_High-Frequency_Data [access: 5 Aug 2021].

[21] B. Yap, A. Mohamed, M. Berry, *Soft Computing in Data Science. Communications in Computer and Information Science*, p. 3.

6.1. Self Organizing Maps and more to find corruption indicators

In case of combating corruption Spanish researches – Felix Lopez-Iturriaga and Iván Pastor Sanz, have developed "an early warning system based on a neural network approach, specifically self-organizing maps, to predict public corruption based on economic and political factors"[22]. The system was to determine the factors that are associated with higher risk of corruption. They have connected various databases from many different areas in the search for correlations. Their research indicate, that "the taxation of real estate, economic growth, the increase in real estate prices, the growing number of deposit institutions and non-financial firms, and the same political party remaining in power for long periods seem to induce public corruption. (Their model) provides different profiles of corruption risk depending on the economic conditions of a region conditional on the timing of the prediction (… and) provides different time frameworks to predict corruption up to 3 years before cases are detected"[23]. The have used unsupervised Self-Organizing maps (SOM) to determine what actually influenced corruption. SOM in principle mimics brain function in pattern recognition. "SOMs have the ability to extract patterns from large data sets without an explicit understanding of the underlying relationships. They convert nonlinear relations among high dimensional data into simple geometric connections among their image points on a low-dimensional display"[24]. So, an unsupervised learning algorithm aim at finding unapparent patterns. In case of corruption, it found correlations of corruption and real estate prices and party staying long in power. Those correlations are far from being obvious. The same scenario could be applied to determine FX valid indicators. Determination may mean finding new ones but also rejecting some of those, which today are unnecessarily considered important and loaded to the described above algorithms and increase their computational complexity. SOMs can work on large databases, so the problem with omitting the emergent features may be also solved, at least less. Even the same set of databases used by the authors to detect the corruption indicators could be used as an initial dataset for FX. Later the results from various correlation research can be combined for example via crossover method to finetune the results. We may acquire surprising results, for example that corruption level is an important indicator in the FX prediction.

There is also an example of successful implementation of a learning algorithm (I do poses technical details of implementation) in combating crime is Brazil.

[22] F. López-Iturriaga, I. Sanz, *Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces. Social Indicators Research*. https://link.springer.com/article/10.1007/s11205-017-1802-2 [access: 5 Aug 2021].

[23] *Ibidem.*

[24] *Ibidem.*

There is "a machine learning application to estimate risk of corrupt behavior among its civil servants. By entering the social security number of a civil servant into the dashboard, the app returns the probability of the person being corrupt, displayed on a simple gauge (…) (where) variables are included in the analysis performed by the tool, such as how the person was employed – by political affiliation or education. Possible criminal records, business and shareholder relations, political affiliations, and level of position are all included in the calculation. A large dataset on convictions of civil servants was used to train the algorithm."[25] Although I was not able to confirm which exact algorithm was employed, a cooperation with its creators may answer this question.

6.2. Active Learning for credit card fraud detection

There is a number of successful implementations of unsupervised learning including credit card fraud detection. The needs of the system seem to be similar as in the case of FX rates indicators searching methods this article is about, since both of them must deal with vague, undetermined input data and indicators. This similarity is visible when Carcillo et al. Claim, that "From a machine learning perspective, however, it is important to keep an adequate balance between exploitation and exploration, i.e. between the short-term needs of providing good alerts to investigators, and the long-term goal of maintaining a high accuracy of the system"[26] . Here it is also difficult to determine the significance or insignificance and label as such of certain activities since the actions are not standard as the behavior of the fraudsters and genuine customers not only changes but overlaps. The authors used Active Learning strategy to manage the situation. There is another similarity with FX market, namely constantly changing conditions. In the article authors advocated the streaming AL version over the pool-based one. The difference is that "In pool-based AL, the algorithm performs queries in the same set of unlabeled points, while in stream-based AL, the set of unlabeled data points is periodically updated. The accuracy of a pool-based AL classifier is expected to grow in time, since more and more labeled data points are used for the Training. This is not always true in the case of the streaming approach, since data received in different periods may differ significantly (e.g. concept drift)."[27]

[25] P. Aarvik, *Artificial Intelligence – a promising anti-corruption tool in development settings?*, https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings [access: 5 Aug 2021].

[26] F. Carcillo, Y. Borgne, O. Caelen, G. Bontempi, *An Assessment of Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection*, https://ieeexplore.ieee.org/document/8259825 [access: 5 Aug 2021].

[27] *Ibidem.*

This preference is also applicable to FX market. The authors performed analyses and also made their code available on Github.

Evaluation of particular methods is not the aim of the article. The aim is to indicate, that currently used FX automated decision and acting methods are working on the basis of selected sets of indicators. The examples of fraud and corruption detection system at least indicate, that it is worth to look in to the issue of selection of relevant set of indicators from the perspective of unsupervised or semi-supervised learning techniques since another area, with analogous issues has been successful. If it proves successful, the results can aid in determining the adequate indicators for building certain models, but also acquired even live and fed as an initial part of a pipeline in ML workflow.

7. Conclusions

This article aimed at stressing the relevance of proper selection of indicators for accurate FX rates predictions and providing an overview of selected methods and tools for it. The list is not complete since both it would exceed the scope of the article and is constantly updated. Yet the aim of the article was to provide the complete list but to show that the idea of using various methodologies to determine the indicators themselves is, to my knowledge, is absent in the domain. Therefore, even if just for the sake of experimentation it should be explored. Especially looking for the fundamental types of principles for building FX prediction models. Creating a set of important, valid indicators and then creating a system based on them could provide better results and then later could create an input for further selection for the described cases appears promising. Even more promising is the scenario, where the indicators are determined life corresponding to changing situation and then delivered live for the system. The initial phase was proven successful in case of other implementations, like corruption indicators detection or fraud identification. Only a question remains how possessing a very accurate, broadly available FX prediction model would influence FX prediction and FX itself.

Bibliography

Lewis M., *Flash Boys: A Wall Street Revolt*, New York – London 2014.

de Laplace P. S., *A Philosophical Essay on Probabilities*, English translation of the French original from 1820 by FW Truscott and FL Emory, 1951.

Hryshko A., Downs T., *System for foreign exchange trading using genetic algorithms and reinforcement learning.*

Koskinen, J. Airas J., Nummelin T., Pekkala T., Starczewski J., *Exploring algorithms for automated FX trading – constructing a hybrid model*, https://www.sciencedirect.com/topics/computer-science/markov-decision-process [access: 5 Aug 2021].

Öztürk M., Toroslu I., Fidan G., *Heuristic based trading system on Forex data using technical indicator rules*, https://www.sciencedirect.com/science/article/abs/pii/S1568494616300369 [access: 5 Aug 2021].

Watkins, C., *Learning with delayed rewards*, https://www.researchgate.net/publication/33784417_Learning_From_Delayed_Rewards [access: 5 Aug 2021].

Moody, J., Lizhong W., *Optimization of trading systems and portfolios*, https://ieeexplore.ieee.org/document/618952 [access: 5 Aug 2021].

Moody J., Saffell, M., *Reinforcement Learning for Trading,* http://papers.nips.cc/paper/1551-reinforcement-learning-for-trading.pdf Access [access: 5 Aug 2021].

Huang W., Lai K., Nakamori Y, Wang S., *Forecasting Foreign Exchange Rates With Artificial Neural Networks: A Review*, https://www.researchgate.net/publication/220385306_Forecasting_Foreign_Exchange_Rates_With_Artificial_Neural_Networks_A_Review [access: 5 Aug 2021].

Dunis, C., Gavridis, M., Harris, A., LeongNacaskul, P., *An application of genetic algorithms to high frequency trading models: a case study*. https://www.researchgate.net/publication/228421645_Different_Methods_to_Clean_Up_Ultra_High-Frequency_Data [access: 5 Aug 2021].

Yap B., Mohamed A., Berry M., *Soft Computing in Data Science. Communications in Computer and Information Science.* https://link.springer.com/book/10.1007/978-981-13-3441-2 [access: 5 Aug 2021].

López-Iturriaga F. J., Sanz I. P. (2017). *Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces*, https://link.springer.com/article/10.1007/s11205-017-1802-2 [access: 5 Aug 2021].

Aarvik P., *Artificial Intelligence – a promising anti-corruption tool in development settings?,* https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings [access: 5 Aug 2021].

Carcillo F., Borgne Y.-A., Caelen O., Bontempi G., *An Assessment of Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection*, https://ieeexplore.ieee.org/document/8259825 [access: 5 Aug 2021].

Shen M., Lee C., Liu H., Chang P., Yang C., *An Effective Hybrid Approach for Forecasting Currency Exchange Rates*. Sustainability 2021, https://www.mdpi.com/2071-1050/13/5/2761 [access: 5 Aug 2021].

Littman M., *Markov Decision Processes*, https://www.sciencedirect.com/science/article/pii/B0080430767006148 [access: 5 Aug 2021].

Rowland Z., Lazaroiu G., Podhorská I., *Use of Neural Networks to Accommodate Seasonal Fluctuations When Equalizing Time Series for the CZK/RMB Exchange Rate* https://www.mdpi.com/2227-9091/9/1/1 [access: 5 Aug 2021].

Sahu K., Nayak S., Behera H., *Forecasting currency exchange rate time series with fireworks-algorithm-based higher order neural network with special attention to training data enrichment* https://journals.agh.edu.pl/csci/article/view/3474 [access: 5 Aug 2021].

Yerashenia N., Bolotov A., Chan D., Pierantoni G., *Semantic Data Pre-Processing for Machine Learning Based Bankruptcy Prediction Computational Model* https://ieeexplore.ieee.org/document/9140238/ [access: 5 Aug 2021].