INNOVATIVE INFOTECHNOLOGIES FOR SCIENCE, BUSINESS AND EDUCATION

Vol. 2(11) 2011

CONTENTS

IITSBE, **Vol. 2(11) 2011**

Investigation of financial market prediction by recurrent neural network

Nijolė Maknickienė[1] [a], Aleksandras Vytautas Rutkauskas[1], Algirdas Maknickas[2]
[1] Faculty of Business Management, Vilnius Gediminas Technical University, Saulėtekio al. 11, Vilnius, Lithuania
[2] Faculty of Fundamental Sciences, Vilnius Gediminas Technical University, Saulėtekio al. 11, Vilnius, Lithuania

Received 17 August 2011, accepted 17 October 2011

Abstract. Recurrent neural networks as fundamentally different neural network from feed-forward architectures was investigated for modelling of non linear behaviour of financial markets. Recurrent neural networks could be configured with the correct choice of parameters such as the number of neurons, the number of epochs, the amount of data and their relationship with the training data for predictions of financial markets. By exploring of learning and forecasting of the recurrent neural networks is observed the same effect: better learning, which often is described by the root mean square error does not guarantee a better prediction. There are such a recurrent neural networks settings where the best results of non linear time series forecasting could be obtained. New method of orthogonal input data was proposed, which improve process of *EVOLINO* RNN learning and forecasting.

Citations: Nijolė Maknickienė, Aleksandras Vytautas Rutkauskas, Algirdas Maknickas. Investigation of Financial Market Prediction by Recurrent Neural Network – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 3-8.

Keywords: Financial Forecasting and Simulation; Time-Series Models; Dynamic Quantile Regressions; Dynamic Treatment Models; Forecasting Models; Simulation Methods; Neural Networks and Related Topics; Recurrent Neural Networks; *EVOLINO* learning algorithm; Non linear time series; Orthogonal inputs; Prediction of financial markets.
JEL: G17; C32; C53; C45.
Short title: Investigation of financial market prediction by RNN.

Introduction

Modeling of non linear processes is actual in two aspects. Realistic models of their history helps to understand the inner structure of nonlinearity. On the other hand, correct understanding of nonlinearity improves the prediction of these processes. For this purpose researchers use the most common Mackey-Glass system [1] as standards tester of non linear processes. The best recognized tools for the finance currency markets is in the last decade neural networks [2-4] or by Reinforcement-Learning Agents [5-6]. Intensive researches of financial market data by neural networks shows that the best learning stage of neural networks does not always lead to correct forecasting.

Financial markets could be explained by using profitability, reliability or risk investment models and analysis methods. Opposite to statistical analysis there could be sophisticated reinforcement learning agents [6] or neural networks [2-4]. The best suited neural networks for the recursive nonlinearity are *Recurrent Neural Networks* (RNN). behaviour of time series in financial, stock or currency markets are influenced by psychology of trades and are strongly non linear and hardly predictable. Using the RNN in modeling of financial time series is based on founding an acceptable learning model for RNN's. RNN's are fundamentally different from feed-forward architectures in the sense that they not only operate on an input space but also on an internal state space [7-8]. For the better improvement of RNN learning the *EVOLINO* algorithm [1] could be selected because it very clearly shows training and validation of the recurrent neural network for non linear data inputs.

The goal of present works is to understand how RNN works for modeling and prediction of the financial markets, their behavioural analysis and paying attention to the acceptance of the chosen method for the anticipation. Author of Ref. [9] was exploring the human mind distinguished reproductive thinking, which only echoes the familiar issues, and productive thinking, that creates something new. In order to solve a specific task certain knowledge is needed.

Meanwhile, not everybody, who has the knowledge required for the task, is able to use it productively. There is no direct link between the past experience and new (productive) thinking. By observing the RNN learning and forecasting the same effect is observed - better learning, which is described by the *Root Mean Square Error* (RMSE), does not guarantee a better prediction.

The aims of this paper are to find the best conditions, where *EVOLINO* RNN becomes a good instrument of financial markets prediction. It will be investigated collection of RNN parameters like number of epochs, number of neurons to achieve strong learning of RNN and good prediction of financial markets. The input orthogonalization method is proposed attaining this goal.

[a]Corresponding author, cell: +370 (687) 25206; email: *nijole.maknickiene@vgtu.lt*

1. *EVOLINO* Learning Algorithm. Description

Neural networks aid to monitor of the non linear processes in the learning activity. The comparison of various methods evaluates neural networks learning algorithms of non linear processes and increase their prediction accuracy. Schmidhuber et al. [10] introduced a general framework of sequence learning algorithm *EVOlution of recurrent systems with LINear Outputs (EVOLINO)* [1]. *EVOLINO* uses evolution to discover good RNN hidden node weights, while using the methods such as linear regression or quadratic programming to compute optimal linear mappings from the hidden state to the output.

When quadratic programming is used to maximize the margin, it is impossible to obtain the first evolutionary recurrent support vector machines. *EVOLINO*-based Long Short-Term Memory (LSTM) can solve tasks that Echo State nets cannot [1]. There was introduced a new class of recurrent, truly sequential SVM-like devices with internal adaptive states, trained by a novel method called *EVOlution of systems with KErnel-based outputs (EVOKE)*, an instance of the recent *EVOLINO* class of methods.

EVOKE evolves recurrent neural networks to detect and represent temporal dependencies while using quadratic programming/support vector regression and pseudo-inverse regression. *EVOKE* is the first SVM-based mechanism which knows how to classify a context-sensitive language. It also outperforms recent state-of-the-art gradient-based *Recurrent Neural Networks* (RNNs) on various time series prediction tasks. RNN learning is used for context-sensitive languages recognition and is a difficult and often increasing problem for standard RNNs, because it requires unlimited memory resources.

For these array of problems investigated by authors of Refs. [1, 10-12], *EVOLINO* based LSTM learns in approximately 3 min on average and it is able to generalize substantially better that gradient-based LSTM. With *EVOLINO* it makes impossible to learn functions composed of multiple superimposed oscillators such as double sine and triple sine. Investigated network reached good learning and still makes very accurate predictions [1, 10-12]. The Mackey-Glass system is a standard benchmark for non linear time series prediction. Authors of Ref. [1] show deviation between the curves of *EVOLINO* generated and Mackey-glass system. *EVOLINO* is capable of making precise (0.0019) prediction in tasks like the Mackey-Glass benchmark.

The block diagram of *EVOLINO* recurrent neural network is shown in Fig 1. *EVOLINO* RNN forms LSTM network with $N = 4n$ memory cells, where N is total amount of neurons and n is amount of memory cells. The genetic evolution algorithm is applied to each quartet of memory cells separately. The cell has an internal state S together with a forget gate (G_F) that determines how much the state is attenuated at each time step. The input gate (G_I) controls access to the cell by the external inputs that are summed into the Σ unit, and the output gate (G_O) controls when and how much the cell fires.

Dark blue nodes represent the multiplication function and the linear regression Moore-Penrose pseudo-inverse method used to compute the output (light blue circle) [13]. The detail description of *EVOLINO* RNN algorithm could be found in Ref. [1], [10].

2. Inputs of Recurrent Neural Network

Authors of Ref. [14] analyzed what inputs should be chosen to receive the best models and got the inputs allowing a better prediction. They have found, that it is common to use the orthogonal inputs, where orthogonality of inputs is equivalent to orthogonality of n dimension vectors. The orthogonality of vectors is result of following inner product for two vectors f and g:

$$\langle f, g\rangle_w = \sum_n f(n) * g(n) * w(n) \tag{1}$$

where $w(n)$ is a non-negative weight vector of the definition of inner product. These vectors are orthogonal if above described inner product is zero:

$$\sum_n f(n) * g(n) * w(n) = 0 \tag{2}$$

This method was selected for finding the best inputs in *EVOLINO* learning process too. The most tools of financial prediction are used for searching dependences between time series of financial indicators or similar patterns in these time series. For this purpose the time series orthogonalization were exploit as follows

$$|\sum_n f(n) * g(n)\ | = \varepsilon. \tag{3}$$

where absolute value of scalar multiplication of vectors ε describes degree of orthogonality, because true orthogonality (2) could not be reached for time series of financial markets and non-negative weight vector is $w(n) = 1$. Prediction of one time series output were obtained by the two most orthogonal time series inputs. The influence of data orthogonality were investigated in the range of $\varepsilon \in [0.00001 \div 0.001]$.

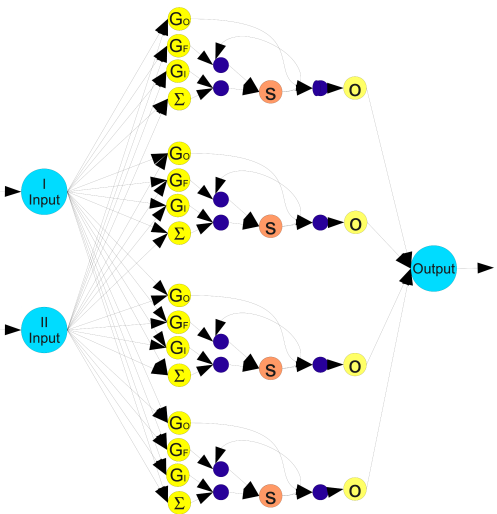

Fig 1. LSTM network with four memory cells.

3. Reproductive and productive learning

Behaviour of human brains could be divided into a productive and reproductive thinking. Reproductive thinking only echoes acquired knowledge and productive thinking creates something new. These brain processes are not straightforward. RNN algorithm ability to learn non linear process is measured and evaluated by RMSE aid. By studying non linear processes, such as a stock or currency markets or fluctuations in solar activity, or others, the RNN prediction is very important research in today areas. It could be distinguished into two major aspects of forecasting: i) how many correct data points of the process could be predicted; ii) how many correct directions of the process could be predicted. If we set itself the objective of accurately predicting of values of the non linear process, we will be facing the problem, as the average value of deviation is acceptable. For such processes, as the shares of a stock or currency market, the prediction of the direction is sufficient for the making the reasonable decision of the future investment. This work will attempt to prove or reject that. The proving or rejection will be confirmed in the investigation of RNN *EVOLINO* algorithm with finding of parameters where the prediction is the best.

4. Criteria learning and predictions

For investigation of *Root Mean Square Error* (RMSE) learning and prediction processes we used program framework [15] adopted for multiple inputs. We selected two parameters - RMSE and correlation for comparison of learned and predicted time series. RMSE is often used in RNN as a learning criteria. Learning of RNN means tests of trained neural network, where data for tests were used after wash out of 1/3 start data of training data. Observation of good RMSE result for learning some times do not imply good forecasting. For this purpose we use correlation coefficient too. As it was mentioned above, moving direction of stock shares or currency ratios is more important than the prediction of their exact values.

The correlation coefficient is located in range $[(-1) \div (+1)]$. A value of $(+1)$ implies that a linear equation describes the relationship between data ranges and predictive ranges perfectly, with all data points lying on a line in which Y increases as X increases. A value of (-1) implies that all data points lie on a line in which Y decreases as X increases.

5. Results and discussion

5.1. Selection of right numbers of epochs

Investigating some certain phenomena of artificial neural networks to work as memory structure or as a predictor is very important to clarify the behaviour of the RNN.

Table 1. Dependence of test RMSE of learning from the number of epochs with data of *USD/JPY* and Gold

Number of epochs	RMSE of learned RNN
16	-0.008298
32	-0.007691
64	-0.007232
72	-0.006012
76	-0.002237
80	-0.002664
120	-0.002270
164	-0.001667
172	-0.001652
188	-0.001523
200	-0.001824
220	-0.001671

It is very important to have not only a well-selected input data and training ranges, but also a good selection of epochs and the number of neural network. Epochs number describes the number of times when NN data are processed, and may wrongly appear that the higher the number of epochs leads to the better learning and prediction. Finally, it was studied the RMSE dependence on the number of epochs, taking the familiar orthogonal data ranges of currency market *USD/JPY* with using of *XAU/USD* as a additional input for improving of convergence.

The obtained results of learning RMSE are taken in Table 1. Results shows that a small number of epochs does not provide RNN learning. Only 76 epochs starts to learn RNN. RMSE dependence on epochs shows that after reached of 164 epochs learning is stabilizing and further increasing of the epochs does not make sense.

5.2. The importance of the number of neurons

The variance of the number of neurons is very important neural networks parameter in the RNN learning. At first glance it may seem that the more neurons are used the better the prediction result will be. But a large amount of neurons took more calculation time and it is important too. Therefore, it is necessary to find optimal number of neurons, which are able to learn and predict data of time series. The results of dependence of RMSE and correlation coefficient from the number of neurons were obtained with of data 85 points of *USD/JPY* with additional input of *XAU/USD*.

The studying the dependences of the number of neurons on learning RMSE could be seen that learning slowly increases, when the number of neurons increases from 16 to 64 neurons. Starting from 68 neurons learning suddenly increases 26 times and the behaviour of RNN learning becomes excellent. Investigation of the dependences the number of neurons on the RMSE and correlation coefficients of prediction could be shown in Table 2.

The three areas of distinct neural networks amount have been found.

Table 2. Dependence of RMSE and correlation coefficient from the number of neurons.

Number of neurons	RMSE of learning	RMSE of prediction	Correlation of prediction
16	-0.009911	1.207037	-0.275067
20	-0.010139	0.269538	-0.334100
24	-0.006452	0.132038	0.000400
28	-0.005201	0.658168	-0.193200
32	-0.004240	0.229054	0.053900
36	-0.003299	0.478200	0.011100
40	-0.002659	0.190520	0.035680
44	-0.002075	0.403500	0.184900
48	-0.001483	0.419292	-0.12925
52	-0.000941	0.258178	0.399926
56	-0.000354	0.363497	0.157300
60	-0.000110	0.560873	-0.042367
64	-0.005167	0.945073	-0.015375
68	-1.600e-29	0.627907	0.068900
72	-8.580e-30	0.757063	-0.051500
76	-1.350e-29	0.533008	-0.383733
80	-5.587e-30	0.312393	-0.224267
84	-7.252e-30	0.665393	-0.272200
88	-1.450e-30	0.481330	-0.227167
100	-2.300e-31	0.882620	-0.292600

The first area is for numbers of neurons from 16 to 40 where averages of correlation coefficients are in interval $[-1 \div 0.1]$. This proves that there are not enough neurons in RNN to learn and to predict. The second area for number of neurons is from 52 to 56 where averages of correlation coefficients are in interval $[0.1 \div 1]$, this proves that RNN try to learn and predict the data. All values of correlation coefficient in this area are in interval $[0 \div 1]$, this proves that RNN predict directions of *USD/JPY* very well.

Third area for numbers of neurons is from 60 to 100 where averages of correlation coefficients fitt interval $[(-1) \div 0.1]$, this proves that increasing of the number of neurons improve learning and RMSE of learning, but suddenly RNN stop to predict. Correlation coefficients versus amount of neurons are presented in Fig. 2. Presented curves shows, that a zone of amount of neurons exists where maximum correlation could be achieved.

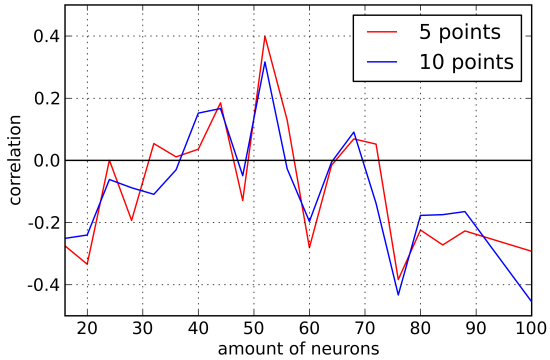

Fig 2. Correlation coefficients versus amount of neurons

The similar results are obtained for five and for ten points prediction.

5.3. Variation of data amount

The last stage of investigation was the variance of the input data size. It was important to know how many days are sufficient to monitor the financial or foreign exchange market in order to obtain reliable forecasting using RNN. In this purpose dependence the number of data on RMSE of learning, RMSE of predicting and correlation of prediction were obtained. Study of dependences the number of data to the RMSE and correlation coefficients and finding of suitable, predictable RNN shows that the RNN behave in the same way as in previous investigations. There are three distinct neural networks areas in the number of data: under learned, best learned and over learned. Dependence the number of data on learning and prediction RMSE and correlation coefficients for *USD/JPY* and *XAU/USD* inputs and *USD/JPY* output is presented in Table 3. Three kinds of behaviour of learning and prediction are given in Fig. 3.

The first area could be separated in which there is not enough data for RNN learning and prediction. The second area of numbers of input data is the best area for RNN learning and prediction of input data. The third area showed that increasing of the number of data improves learning and RMSE of learning, but RNN stop to predict and the further increase of number of neurons do not imply better prediction.

All three studies have shown that the RNN prediction could be obtained when the neural network parameters such as epochs, number of neurons and the number of data are in a certain range.

Table 3. Dependence on learning and prediction RMSE and correlation coefficients from the number of data

Number of data	Number of neurons	RMSE of learning	RMSE of prediction	Correlation of prediction
50	36	-0.000054	0.09935	0.4528
57	36	-0.000415	0.29459	0.5328
65	36	-0.001570	0.2077	0.7571
70	36	-0.002385	0.29763	0.9103
76	36	-0.002808	0.07895	0.1006
85	36	-0.001714	0.1286	0.8726
85	64	-0.007561	0.09726	0.3360
90	64	-0.001452	0.1564	-0.0286
95	64	-0.000714	0.18421	-0.7302
100	64	-0.000990	0.14608	-0.7320
105	64	-0.001245	0.3206	0.6198
110	64	-0.001777	0.23879	0.7114
115	64	-0.002061	0.2093	-0.2290
120	64	-0.004170	0.25555	-0.5788
125	64	-0.003538	0.16155	-0.7229
130	64	-0.003216	0.14268	-0.8211

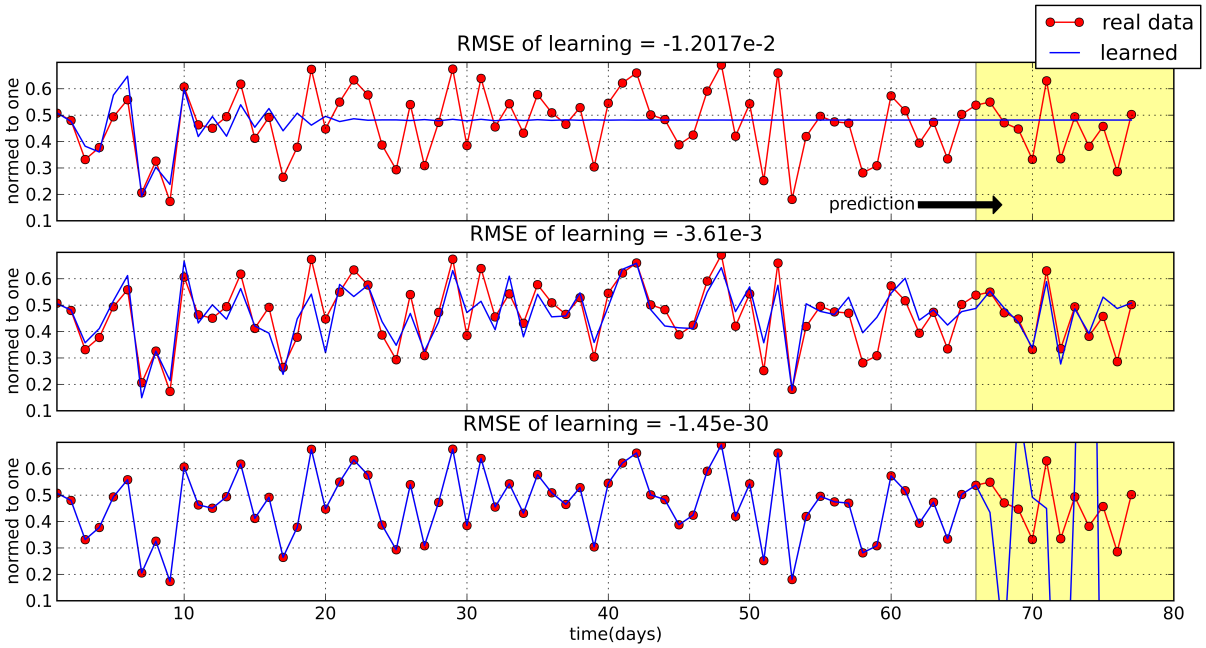

Fig 3. RNN learning and prediction. Three types of tests.

Our studies have shown the most 164 epochs is enough. Number of neurons must be in interval $[52 \div 56]$ and number of data in intervals $[80 \div 85]$ or $[105 \div 110]$. The averaged correlation coefficient of forecasting for this range of initial parameters of RNN were reached value of 0.32 and, in separate studies, even 0.9579.

6. Conclusions

The aim of the presented work was finding of the best conditions where the RNN makes the best prediction of currency markets. It was investigated that the prediction of the evolution of the *USD/JPY* exchange daily rates for 15 March 2010 in 5 following days. Data were collected from 1 Januar 2009 till 15 March 2010. *USD/JPY* exchange rates trained for the same period of *XAU/USD* data inputs. The obtained results show that.

1. The lowest values of orthogonality degree description coefficient ε improve stability of RNN learning and prediction of investigated non linear time series. The confirmation of quantitative dependences needs future investigation.
2. The learning and prediction of RNN, like the human brain productive and reproductive thinking, are independent and different. The better RMSE of learning do not guarantee the better achieving of prediction.
3. Combinations of parameters of RNN such as the number of epochs, data and neurons amount, determine different behaviour of learning and prediction. weak learning without prediction; strong learning with prediction; excellent learning without prediction.
4. The investigation of financial data gives such group of parameters of RNN in *EVOLINO* algorithm, where RNN predict directions and values of a currency market. The group of RNN parameters for given data was found where the average of correlation coefficient of forecasting reaches maximum 0.938 and has value equal to 0.400.

References

1. J. Schmidhuber, D. Wierstra, and F. Gomez. Evolution: Hybrid neuroevolution / optimal linear search for sequence learning. – Proceedings of the 19th International Joint Conference on Artificial Intelligence (2005) 466–477.
2. M. Adya and F. Collopy. How effective are neural networks at forecasting and prediction. A review and evolution. – *Journal of Forecasting* 17 (1998) 481–495.
3. P. McNelis. Neural Networks in Finance: Gaining Predictive Edge in the Market. – Elsevier Academic press, 2005.
4. D. Plikynas. Explaining international investment patterns: a neural network approach. – *Information technology and control* 34(1) (2005) 42–50.
5. T. Ramanauskas and A. Rutkauskas. Empirical version of an artifical stock market model – *Monetary Studies* 1 (2009) 5–26.
6. A. Rutkauskas and T. Ramanauskas. Building an artificial stock market populated by reinforcement-learning agents. – *Journal of Business Economics and Management* 4 (2009) 329–341.
7. J. L. Elman. Finding structure in time. – *Cognitive Science* 14 (1990) 179–211.

8. M. Boden. A guide to recurrent neural networks and backpropagation. Technical report, DALLAS project. Report from the NUTEK supported project AIS-8, SICS.Holst: Application of Data Analysis with Learning Systems, 2001.
9. E. Rimkutė. Mąstymas ir kalba (in lith.). – Vilnius: University press, 2007.
10. J. Schmidhuber, M. Gagliolo, D. Wierstra, and F. Gomez. *EVOLINO* for recurrent support vector machines. In: ESANN'2006 proceedings - European Symposium on Artificial Neural Networks, Apr. 2006, ISBN 2-930307-06-4 (2006) 593–598.
11. B. Schrauwen, D. Verstraeten, and J. Campenhout. An overview of reservoir computing: theory, applications and implementations. In: Symposium on Artificial Neural Networks, ISBN 2-930307-07-2. (2007) 471–482.
12. D. Wierstra, F. Gomez, and J. Schmidhuber. Modelling systems with internal state using *EVOLINO*. In: Conference on genetic and evolutionary computation GECCO. (2005) 1795–1802.
13. R. Penrose. A generalized inverse for matrices. – In: *Proceedings of the Cambridge Philosophy Society* 51 (1955) 406–413.
14. V. Rutkauskas, N. Maknickienė, and A. Maknickas. Approximation of dji, nasdaq and gold time series with *EVOLINO* neural networks. – 6th International Scientific Conference *Lithuania Business and Management*, May 13-14, 2010,Vilnius. 1 (2010) 170–175.
15. D. Goodman and R. Brette. Brian: a simulator for spiking neural networks in python. – *Frontiers in Neuroinformatics* 2(5) (2008).

Modeling of business processes. 1. Overview of models and metamodels

Sigitas Vilkelis [a]
Department of Informatics, Faculty of Mathematics and Informatics
Vilnius University, Naugarduko 9, Vilnius, Lithuania

Received 28 June 2011, accepted 10 September 2011

Abstract. Different aspects of business process modeling are observed and table structure describing business processes was formalized. Methods and categories of modeling, as well as terminology, existing techniques and tools for *Model Driven Architecture* (MDA) are revised and possibilities of their application in transformation of business process models are evaluated.

Citations: Sigitas Vilkelis. Modeling of Business Processes. 1. Overview of Models and Metamodels – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 9-15.

Keywords: Model Driven Architecture; MDA.
JEL: C88; M15.
Short title: Overview of models - 1.

Introduction

Nowadays, under the influence of IT on business, it is difficult to imagine a successful but completely not computerized business. The influence of IT on computerization of specific business processes is essential and has many forms - from digital data storage to automatization of difficult multi-user business processes. Creation of business models is very important in keeping up with the advancing technologies. They are necessary for a change of system platform. In order to create a successful model, specific business terminology has to be known along with the software terminology. The resulting model is a combination of two different subjects: the business terminology and the abilities of business process modeling tools.

This is the reason why business models are usually created by two groups of people: the ones employed in a specific area who know the details well and programmers creating the software system to implementing the business model. Business analysts receive the requirements of initial business processes from customers and pass them to programmers.

Althrough the phase of requirement clarification employs several people, a fixed business process description structure is required. It is needed to be able to look for similar processes in the same structure when creating models for other customers as well as enabling formal documentation and creation of requirements that should be met.

The technologies have been advancing rapidly, but customers still use simple tables, filling the forms by type or even by hand. Models created this way have much higher error probability and it takes additional time converting them to modern standards processed by a computer. Switching to structures defined by the modern standards would reduce the workload for everyone - from analysts, communicating with customers, to programmers, doing all the software package creation for the same customer. Due to the customer's habit of using tables, there is a need for special instruments designed to convert these tables to processes understandable by computers.

This work is devoted to the literature review containing business process modeling and its application of architecture related to these models. Methods and categories of modeling as well as terminology, existing techniques and tools for *Model Driven Architecture* (MDA) are revised and possibilities of their application in transformation of business process models are evaluated.

1. General formulation of a task

According to the experience of the author, the following problems occur when using table modeling.

1. A process modeled using tables will not avoid making a lot of errors, until actual programming takes place.
2. Such process may not be added to running business management systems.
3. Logical business process errors are difficult to notice in a spreadsheet description.

The first problem may be solved in three steps. First of all,

[a]Corresponding author, email: *sigitasv@gmail.com*

the structure of a table should be defined. Secondly, rules should be set to define the formal table structure and finally, these rules should be written in formal form.

The solution to second problem may be expressed as a creation of a process to transform table-based business processes to ones using standard notation. This enables easier manipulation with these processes in management of software systems. The third problem may be solved by applying the same method as in the second problem. The transformed process may be represented graphically (as two-dimensional or three-dimensional distribution), which would help to avoid logical errors.

These are the general problems occurring in process modeling. By clarifying these problems, the following steps are pointed out for this work.

A. Define the business processes described in the tables.
B. Create a process of transformation of table-based business process to formal business process notation.
C. Create a structure of rules and define the rules that processes described in tables should follow.
D. Demonstrate the process of rules and transformation inspection by releasing a prototype of business process transformation and verification.

In order to achieve the A-task, partial sub-tasks were done.

A1. Analysis of existing business process notations was done and a common set of elements was taken from them.
A2. According to the resulting list, formal structure was made and fields of data as well as data types within fields were described.

The steps to complete the B-task are presented below.

B1. Architectural ideas based on models for transformation of table-based business processes to formal business process notation were revised and applied.
B2. Rules of transformation were defined.

C-task was completed by performing the following steps.

C1. A set of rules to be used by table-based business process was defined.
C2. This set of rules was applied to formal table-based business processes.

Actions taken to accomplish the D-task were described in three positions.

D1. According to the business modeling at a workplace of the author and analysis of References, entirety of technological tools for practical solution was made.
D2. A tool from a selected technological environment was chosen to ensure the transformation of table-based business process to formal business process notation.
D3. The selected tool was extended with a rule inspection.

The specific steps and their execution are described in following chapters.

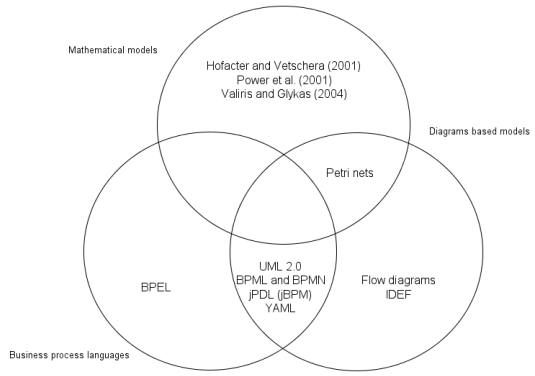

Fig. 1. Classification scheme of the business process modeling. Adapted according to Ref. [1].

2. Methods and notations. Review

Business processes may be modeled using various methods, such as mathematical, chart plotting tools or languages defining business processes. Fig. 1 represents the classification scheme of a business process modeling.

As we can see in the Fig. 1, strict technique separation into three categories is absent, because, some of them have common features. All categories are reviewed below.

2.1. Categories of business process modeling

Mathematical Methods. Mathematical methods are based on strict formalism which results in models which are completely accurate and satisfying all requirements. These models have fewer errors and may be verified using formal methods. Nevertheless, it is inconvenient to create business models using mathematical methods, because advanced mathematical, logical and field-specific knowledge is required.

Chart-based models. The biggest advantage over mathematical methods to chart methods is that they are defined not in mathematical formula, but in graphical charts. This is very important feature, when there is a need to present it to a customer. The business processes are less formal this way, but much more popular.

Languages of business processes. The most popular models are based on languages of business processes. The popularity is based on the applicability of it - all languages can be represented graphically and have their unique document saving format. . Graphical representation is important for understanding the process and the possibility to save them provides a possibility to exchange the documents between different tools.

2.2. Description of business process modeling

Modeling of business processes could be performed using specific tools (created for business process modeling only),

although there are some tools that are general modeling tools with extensions enabling business process modeling. Most of the methods exist as standard tools, such as *Object Management Group* (OMG), *Unified Modeling Language* (UML), *Business Process Modeling Notation* (BPMN), *XML Process Definition Language* (XPDL) [2].

BPEL4WS. Modeling standard BPEL4WS was created by unified forces of BEA, IBM, Microsoft and other companies. The purpose of it is describing of processes provided on Internet and enables creation of difficult business processes, by combining several independent actions to one set of jobs. Such combination is suitable for business process modeling and it is based on *Web Service Definition Language* (WSDL), so different web services exchange XML documents. Service oriented (SO) business process modeling has its own advantages over methods using strict process arrangement. They are: flexibility and easiness of changes [3].

BPMN. BPMN is a standard created by *Business Process Management Initiative* (BPMI), describing Business Process Diagram (BPD) based on diagram creation technique used in graphical representation of business process models. This standard is very popular amongst analysts as well as programmers, because it puts the client needs in a simple way, but it has only the graphical notation, lacking formal description of business process. BPMN documentation [4] also lacks definition of how should the graphical data be stored in a way that would be understandable by computer. This creates an uncertainty of a format used, because every tool uses its own data format (incompatibility could occur). Although, there are some tools that use XPDL's XML documents [5].

There are several important notation problems of BPMN explored in Ref. [6]. This notation has nothing in common with representation of user interface of business process in program system and there is no connection between the business process and the modeled field. The authors of Ref. [6] found a solution to both of these problems. Using rules of transformation, a problem of transforming BPMN diagrams into YAWL diagrams is analyzed in Refs. [7-8]. Working tool is able to perform such transformation, which also shows, that BPMN notation described formally may be transformed into any other formal structure.

UML. UML is one of the most popular program systems and business process modeling languages. It defines many different diagrams made for objective system modeling, and diagrams such as activity enable modeling of fully-fledged business process. Business processes may be defined as an oriented graph, made of peaks and bows using UML. Peaks represent the performed single or combined activity. Peaks may be used for execution control as well. Start and end peaks are assigned to execution control peaks, marking the beginning and the end of a process. Connection and branching in UML may be used to model separation or addition of several parallel processes [2].

UML sequence diagram can be used to show exchanging the information between people participating in business process. Extended activity diagram is suitable for creation of a model describing communication between people [9]. Modeling of variability using UML is presented in Ref. [10]. Common fields where changes are usually made are presented in this paper, these are the changes during the business process, data transmission sequence or activities of business process. It is said, that the description of business process should not go into details, but instead represents only the general sequence of actions making space for actions that could possibly change over time. Nevertheless, this method can only be theoretical, because it is essential to specify common input and output points, sequence of activities, even the exact processed data in practical business process.

The information provided by Ref. [11] describing the transformation of UML business process to XPDL business processes is very beneficial. The general idea is to fill in the missing data in UML diagrams and transform them by using XSLT to XPDL. In order to have XPDL documents fully structured, UML diagrams are filled with extra stereotypes.

XPDL. XPDL is a language of business process description introduced by *Workflow Management Coalition* (WfMC) to define a general data exchange form and supporting moving of different process description between different tools. The purpose of XPDL 2.0 is description of business processes presented in BPMN graphical notation.

Petri net. *Petri net* is a way of business process modeling using mathematical methods and graphical imaging. It consists of places, transitions and arcs. Places may be marked and moved to other places by following the rules. It is very convenient to describe and analyze parallel, asynchronous or distributed systems. As a graphical tool, Petri net can be used to represent graphical connections similar to sequence diagrams. Simulation of dynamic and parallel systems is also performed by using bookmarks in these Petri nets.

The example of business process modeling is presented in Ref. [12]: the basic of modeling is based on two components - activities and resources. Specific resources, such as human labor, specific data or even Internet services, are required for these activities. The biggest benefit of using Petri net is the scalability of models: using several layer modeling, even the smallest processes may be modeled and combined to larger ones, thus creating a clear and detailed business model.

Another similar sample of modeling is described in Ref. [13]. The main difference between the previous one is the usage of several Petri nets: ontological, based on abstract understanding of business process, concept, introducing the

business transaction term and system interfaces and functional net describing interfaces, services and data streams. This type of model enables the evaluation of a business process from three different angles, which proves useful when there is a need to confirm the logic of a process.

IDEF0 and IDEF3. The purpose of *Integration Definition for Functional Modeling* (IDEF0) is functional modeling based on usage of text and graphical markings on organized and systematic models. This increases the understanding of project and integration activities as well as defines the requirements.

Method called *Integrated DEFinition for Process Description Capture* (IDEF3) defines the collection and documentation of processes. It incorporates easy-to-understand priorities and connections between actions.

VPML-S. A new graphical business process modeling language *Service-Oriented Visual Business Process Modeling Language* (VPML-S) was created as a language based on UML, extended with stereotypes [14]. Its purpose is modeling of service-based business processes. The main goal was to create a language that would have a decent graphical notation and would not require specific IT knowledge in order to use it. A business process written in this language is fully compatible with BPEL language, which defines business process as a septum: activities, products, resources, connections, events, attributes, partners. Every part of the septum is strictly defined using mathematical methods. It may be stated that it can be used to model business processes because it supports Internet services, although there are a couple of problems - there may be a lack of support for the language and poor availability of tools enabling modeling in this language, because it is fairly new, written in 2008 and designed for academic purposes.

Existing Activity Diagram. Modeling of a Business Process could be based on *Existing Activity Diagram*. If a business has documentation of its activities as activity diagrams, they may be transformed to business processes as if reusing them [15]. Such modeling of business processes saves time on analysis and documentation of existing business processes. It may also increase the quality of business processes and reduce the error probability. The information on usage of such diagrams should be retrieved at first, in order to know whether they are still valid: for example, if a company kept records of such diagrams for their first year, but discontinued afterwards, there is no use for these diagrams.

JBPM and JPDL. *JBoss Business Process Management* (JBPM) is a management system, filling the gap between analysts and programmers. It is flexible and provides a way of process modeling, suitable for both of these groups.

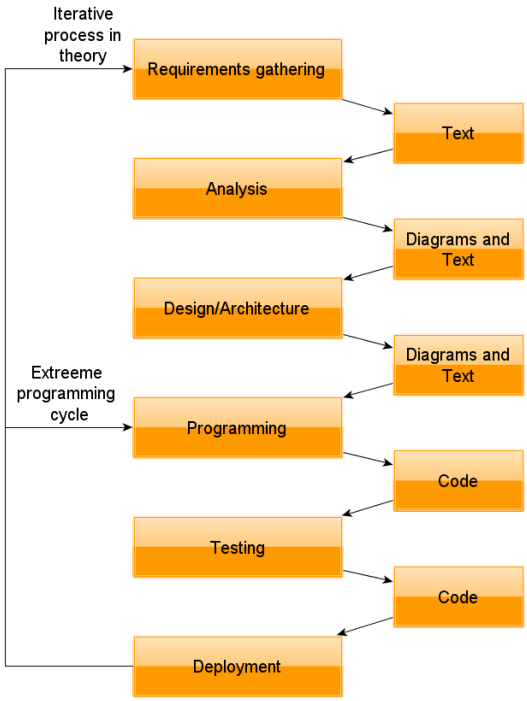

Fig. 2. Traditional cycle of software creation. Adapted according Ref. [16].

Input data for JBPM is presented as descriptions of graphical business processes. The process represents a sequence of actions that are defined as transitions from one activity to another. These graphical diagrams of business processes are the basic way of communication between analysts and programmers.

JBPM as JBPM Process Definition Language is based on *Process Virtual Machine*, which is able to support several languages devoted to the business process definition. JPDL is currently the basic language, created by a business itself. JPDL is a flexible language with extension possibilities, which, according to experience of author, enables easy implementation of JBPM JPDL processes to active systems [16].

3. Transformation technique

3.1. Models and meta-models

The amount of tools and techniques for digitization of businesses has been growing constantly as well as the amount of digitized businesses themselves. One of the systems was based on creation of a model for every step describing it with a required level of detail. This method prevents creation of many documents and allows transformation of the model to a certain software. Such technique was named *Model Driven Software Development* (MDSD). Traditional cycle of software creation is shown in Fig. 2, and extended cycle of *Model Driven Architecture*(MDA) based software creation is shown in Fig. 3. Several software development areas are based on MDA. One of them is called *Model Driven Engineering* - MDE.

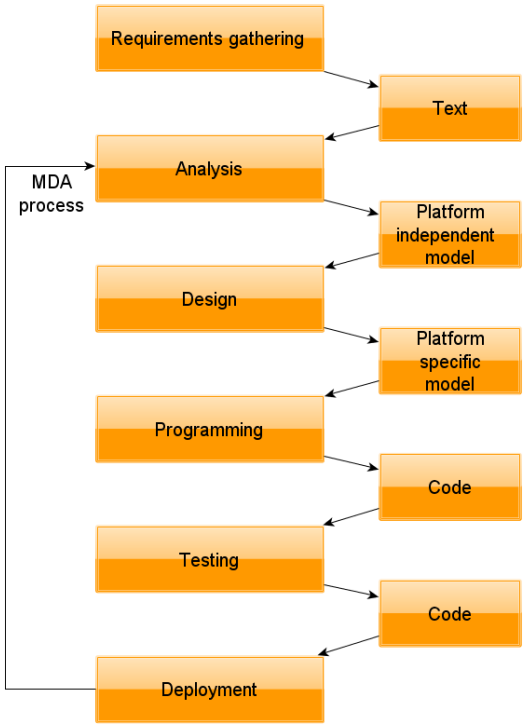

Fig. 3. MDA based software cycle creation. Adapted according Ref. [18].

MDA is developed by OMG group and it specializes not only in software development but in separating logic of business and software from a specific technology in software. A digitized business solution defined by *Platform Independent Model* (PIM) based on UML and other OMG standards with MDA ideas may be implemented in any specific platform using web services, such as .NET, J2EE and others. PIM models define a specific software functionality required by business separately from a software based on specific technology.

That way technological restrictions are avoided and moving to another technological environment freely is encouraged. *Platform Specific Models* (PSM) are derived from PIM models and then transformed to a software supported by specific PSM environment [18-21]. Relation between PIM and PSM is shown in Fig. 4.

One of the general properties of MDA is transformation of models. During one transformation, PIM is joined with additional information and PSM is generated. During another transformation, a realization of software is generated using mapping method. A realization of specific transformation depends on the software system. A kind of transformations exists when models written in PIM language are transformed to models written in PSM language. PIM and PSM meta-models and rules of transformation are defined to enable such process. This transformation is performed between two PIM and PSM models with specific values. Graphical representation of this process is shown in Fig. 5.

Fig. 4. Relation between PIM and PSM.

3.2. Types of model transformations

Model-to-Model (M2M) transformations have been presented previously [17, 19]. There is another way of transformations - the so-called *Model-to-Text* (M2T) transformation. It converts a model to any text: from software code to documents of any format. A lot of various tools can be either commercial or open-source. Both types will be reviewed.

Open-Source Model Transformation Tools. *Kermeta* package was created by INRIA Triskell [20]. It is based on Eclipse platform and the environment is of object-oriented type. The purpose of it is describing and transforming of models and meta-models, as well as their simulations. Kermeta is created as an extension to *Eclipse Modeling Framework* (EMF).

MOFScript is a tool for M2T transformations based on EMF as well. Its purpose is transformation of models and metamodels based on Meta-Object Facility (MOF).

The IBM *Model Transformation Framework* (MTF) - is a tool for describing relations between meta-models in QVT and it is based on EMF as well.

The *ATL Engine* - a language similar to QVT written by INRIA Atlas. It is one of the most important technologies in Eclipse M2M project, created as a bunch of add-ons and it works as a built-in programming language to perform, describe and trace transformations between models [22].

OpenArchitectureWare (oAW) - a flexible framework working along XMI and based on templates.

Generative Model Transformer (GMT) - an Eclipse project for a model transformation technology for Eclipse. Several current tools are a part of GMT: AMW (Model Weaving), Epsilon (Model Merging), MoDisco (Model Discovery), MOFScript (M2T), openArchitectureWare, UMLX (Graphical Transformation), VIATRA2 (Visual Automated Transformations).

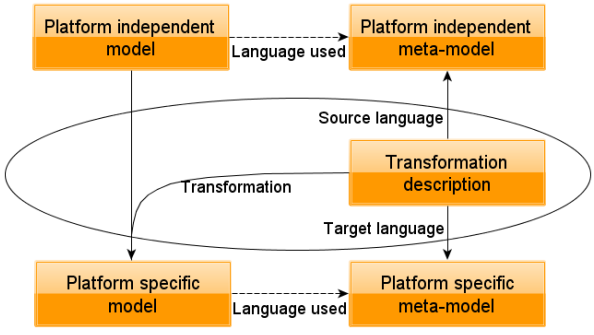

Fig. 5. Process of transformation when using meta-models. Adapted according Ref. [21].

OpenMDX - open-source MDA environment consisting of several tools based on XMI and supporting multi-platform (J2EE, .NET) code generation [21].

Commercial Model Transformation Tools. *ArcStyler* - a commercial MDA tool created by Interactive Objects. It is sold along with MagicDraw UML tool, but it supports other UML tools as well.

Model Component Compiler (MCC) - a commercial product of InferData supporting transformations of M2T to J2EE.

Xactium XMF Mosaic - a tool supporting M2M transformations.

Model-in-Action and MDA - a tool created by Mia software, that supports generation of software and M2M transformations on flexible framework.

MetaEdit+ - built-in environment of modeling and meta-modeling for creation of languages and source generation. It supports XML and SOAP/Webservice transformations for models and meta-models.

MDWorkbench - a tool supporting M2T and M2M transformations accepting any meta-model format as an input. It is based on Eclipse and EMF [22].

Conclusions

Business process modeling of author's workplace was analyzed and table structure describing business processes was formalized. A process of transformation of a table-based business process to a standard description business process notation, based on MDA, was defined. Rules that have to be met by table-based business process were structured and described. Transformation and rule checking were implemented using Eclipse with standard and oAW plugins.

A successful prototype and process definition prove that objectives were completed successfully. The achieved results will improve and speed up work of several people: analysts will be able to check and see graphical business process while filling the business process description table. This will prevent logical errors that occur while creating a business process from separate tasks. Programmers will be able to retrieve the notation which are structured and depicted in specific form of business process. Transformations will be done directly from table-based business process. They will only need to fill some additional information and implement it to a running system. The transformation process using BPMN notation has obvious advantages - business processes described by this notation can be transformed to almost any other business process notation providing almost unlimited expandability for such process. This possibility enables company modeling their processes by tables to adjust their final result to used technologies or specific client requests without changing their initial table-based business process transformation to standard description business process.

Acknowledgments

This work was prepared as a part of the Master Thesis. Author would like to thank Prof. Rimantas Vaicekauskas (Vilnius university) for valuable comments and suggestions.

References

1. Kostas Vergidis, Ashutosh Tiwari, Basim Majeed. Business Process Analysis and Optimization: Beyond Reengineering. – *IEEE* 2007, 14p.
2. Wei Wang, Hongwei Ding, Jin Dong, and Changrui Ren. A Comparison of Business Process Modeling Methods. – *IEEE* 2006, 6p.
3. Corine Cauvet, Gwladys Guzelian. Business Process Modeling: a Service-Oriented Approach. – *IEEE* 2008, 8p.
4. Object Management Group. Business Process Model and Notation (BPMN). Version 1.2, 2009-01-03. – < http://www.omg.org /spec/BPMN/1.2/PDF/ >, accessed 2010-05-03.
5. Michael zur Muehlen, Danny T. Ho. Service Process Innovation: A Case Study of BPMN in Practice. – Proceedings of the 41st Hawaii International Conference on System Sciences, 2008, 10 p.
6. Dagmar Auer, Dirk Draheim, Verena Geist. Extending BPMN with Submit/Responsestyle User Interaction Modeling. – *IEEE* 2009. Conference on Commerce and Enterprise Computing, 2009, 7p.
7. JianHongYE, ShiXin SUN, Wen SONG, LiJie WEN. Formal Semantics of BPMN Process Models using YAWL. – Second International Symposium on Intelligent Information Technology Application, 2008, 8p.
8. JianHong YE, ShiXin SUN, Wen SONG, Lijie Wen. Transformation of BPMN to YAWL. – International Conference on Computer Science and Software Engineering, 2008, 6p.
9. Wang Hongxiu, Wang Gang, Wen Xiaoxian, Gao Guoan. Business Process Modeling for Multi-enterprise Cooperation. – *IEEE* 2007, 4p.
10. Maryam Razavian, Ramtin Khosravi. Modeling Variability in Business Process Models Using UML. – *IEEE* 2008, 6p.
11. Ping Jiang, Quentin Mair, Julian Newman. Using UML to Design Distributed Collaborative Workflows: from UML to XPDL. – *IEEE* 2003, 6p.
12. Hongmei Gou, Biqing Huang, Wenhuang Liu, Shouju Ren, Yu Li. Petri-Net-Based Business Process Modeling for Virtual Enterprises. – *IEEE* 2000, 6p.
13. Xin Wang, Yanchun Zhang, Hao Shi. Scenario-based Petri Net Approach for Collaborative Business Process Modelling. – *IEEE* 2007, 8p.

14. Shaomin Xing, Bosheng Zhou, Tianying Chen. A Service-Oriented Visual Business Process Modeling Language. – *IEEE* 2008, 5p.
15. Chao Yu, Guoqing Wu, Mengting Yuan. Business Process Modeling Based on Workflow Model Reuse. – *IEEE* 2005, 4p.
16. JBoss. JBoss jPDL. – <http://www.jboss.org/jbpm >, accessed 2010-05-01.
17. A. Kleppe, J. Warmer, W. Bast. MIDA Explained. The Model Driven Architecture: Practice and Promise. – Addison-Wesley, 2003, 170 p.
18. OMG. OMG Model Driven Architecture. – <http://www.omg.org/mda/>, accessed 2010-05-01.
19. Eclipse Foundation. Model To Model. – <http://www.eclipse.org/m2m/>, accessed 2010-05-01.
20. Eclipse Foundation. Eclipse Modeling Framework Project. – <http://www.eclipse.org/modeling/emf/>, accessed 2010-05-01.
21. Igor Sacevski, Jadranka Veseli. Introduction to Model Driven Architecture (MDA), Seminar Paper. – 2007-06, 15p.
22. Pau Giner, Victoria Torres, Vicente Pelechano. Bridging the Gap between BPMN and WS-BPEL. M2M Transformations in Practice. – 2006, 14p.

Modeling of business processes. 2. Prototypes of transformation

Sigitas Vilkelis [a]
Department of Informatics, Faculty of Mathematics and Informatics
Vilnius University, Naugarduko 9, Vilnius, Lithuania

Received 28 June 2011, accepted 10 September 2011

Abstract. Working method devoted to formal table-based business process transformation using Eclipse to formal business process notation is presented here. The resulting formal business process notation document may be used as a template in further programming jobs by implementing it to a working system.

Citations: Sigitas Vilkelis. Modeling of Business Processes. 2. Prototypes of Transformation – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 16-21.

Keywords: Model Driven Architecture; MDA; Eclipse.
JEL: C88; M15.
Short title: Prototypes of transformation - 2.

Introduction

Previous publication [1] is aimed to review the different aspects related to transformation between business processes. Methods and categories of modeling, as well as terminology, existing techniques and tools for *Model Driven Architecture* (MDA) were revised [2-4] and possibilities of their application in transformation of business process models are evaluated.

The basics of the business process transformation and requirements of expandability are defined, as well as the input and output data on every step of transformation. A choice of using indirect transformation, by adding a temporary business process notation, instead of direct transformation is explained.

A working method devoted to formal table-based business process transformation using Eclipse to formal business process notation is presented here. The resulting formal business process notation document may be used as a template in further programming jobs by implementing it to a working system.

1. MDA Tools for models

MDA is a new outlook to a software, implementing traditional models by using them as input or output data. The purpose of MDA is replacing traditional diagrams, such as UML [5] or ordinary text by models. This allows access to lower layer models by applying transformations, generating machine code at the lowest layer. All these transformations are defined by specific rules of transformation from one model to another [6].

MDA technique was used in this work to perform a transformation from a business process described in a table to formal business process notation.

According to author's experience, analysts of business systems still use tables instead of standard approved techniques and methods when communicating with their customers to provide easier apprehension throughout computerization of a business. The objective of this work was created to make sure that transformation of structured tables to a digital format understood by computer is possible. This provides several ways to find the solution.

One of them is the transformation of documents described above using XSLT. It is not very convenient, because if the initial table structure changes, which may occur when analyst has to adjust it for different customers, the structure of XSLT transformation has to change too. This may be a problem while working with several customers at once as well, because different XSLT transformations are needed for every separate table.

Another way is to separate variable part of table-based business structure from standard description business process describing the resulting non-variable part. A transformation between table-based business structure and standard description process should be performed by defining metamodels describing it, because standard description process has to always remain unchanged, even if the table-based process changes. The most convenient representation of table-based process structure is a CSV document with a predefined row-column structure.

The standard description business process can be defined using BPMN and a transformation between CSV and BPMN

[a]Corresponding author, email: *sigitasv@gmail.com*

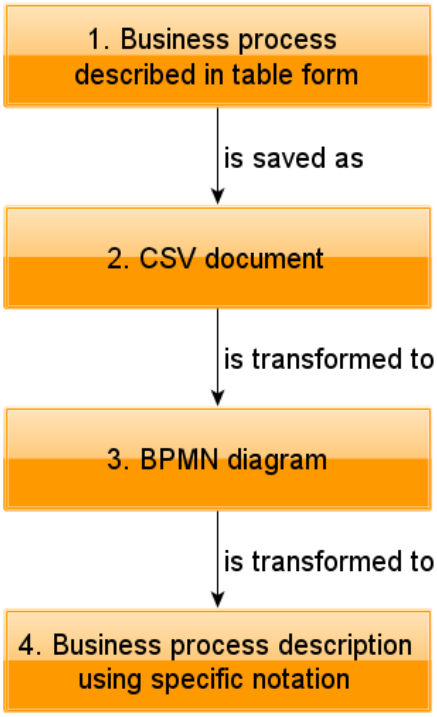

Fig. 1. Transformation of a formal business process table to a specific business process notation.

meta-models may be performed [7].

The BPMN structure may not be applied to a system directly, therefore a transformation to a specific business process notation has to be performed. Fig. 1 represents the transformation process graphically.

A third way to solve the problem of transformations is saving a formal table-based business process as a CSV or any other document format capable of maintaining organized row-column structure. Every section of CSV document is separated by using a comma or semicolon. Several sections constitute a record which takes up one row. A table of business process description in such form may be analyzed and processed by using computer tools.

To avoid scenario presented in first solution, variable and non-variable parts have to be separated. The conversion to specific business process notation is performed after transformation of CSV document to standard business process notation, which is BPMN in this case. As the analysis of literature has shown, business processes in BPMN are much clearer and accessible with many software development tools. The notation itself has a wide choice of available elements ensuring proper conversion to a chosen specific business process notation.

The conversion between CSV and BPMN may be performed by using XSLT transformation as well, although a problem of supporting the platform appears. In order to avoid it, meta-models describing CSV and BPMN business processes should be defined. After that, transformation rules have to be introduced as well. Another part of the process is transformation of BPMN to a specific business process notation. They should be chosen according to the experience of maintenance personnel and current software running in a business they are going to be used on and to ease the implementation. The meta-model of specific business process notation, as well as the transformation rules have to be defined.

The process described above, enables the following:

i) the meta-models and transformation rules between BPMN and specific business process notations may remain constant when the business processes defined in CSV change;
ii) changing the transformation rules independently of CSV, BPMN or specific business process notation meta-models;
iii) transformation of business process defined in BPMN to any specific business process notation, so changing platforms may be performed keeping old and working business processes by transforming them to the new notation.

As we can see, the resulting specific business process notation is ready to implement along the modeling tools and technologies used in business, even if it is not completely full because of specifics of different notations.

Still, the resulting notation may be a decent fundament for further implementations and it minimizes the amount of human labor for conversion of a table-based business process to a specific business process notation. A prototype resembling this transformation will be presented in next chapter.

2. A Prototype of Business Process Transformation

A prototype of transformation of formal table-based business process to a standard description business process was released by using a plentiful list of literature and the theoretical part revised above.

2.1. General Process of Transformation

A decision to perform an experiment of transformation of CSV document to a specific business process notation using a prototype was made. The specific business process notation was selected to be a notation based on JBPM JPDL standard used in author's workplace. The graphic representation of the bonds between models and data are shown in Fig. 2.

It is based on M2M transformations which are described by standard and the oAW plugins of Eclipse [8-9]. The crucial transformation is between CSV Ecore model and the Ecore model of general BPMN elements used for other standards. When the requirements of the prototype were known, open-source Eclipse environment was chosen, supplemented with standard and oAW plugins:

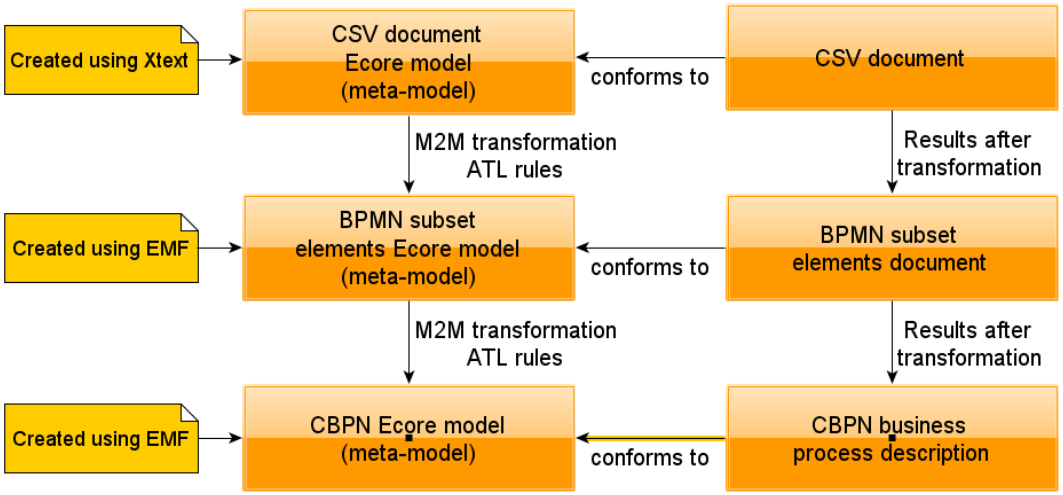

Fig. 2. Bonds between models and data used in transformation.

i) EMF/Ecore was used for description and securing of meta-models;
ii) ATL - a plugin supporting transformation language enabling creation and execution of transformation rules [10];
iii) oAW Check - a tool for description of rules of meta-models defined by Ecore construction, enabling the confirmation whether a specific meta-model follows the rules.

The graphical environment of technology is shown in Fig. 3.

At first, the structure of CSV document was analyzed and *Domain Specific Language* (DSL) was defined using the Xtext plugin for Eclipse. The DSL was used to get an Ecore model, describing the structure of CSV document in terms of Eclipse EMF framework. Ecore is a meta-model of EMF framework, supporting saving of models in XMI format. XMI is a standard for XML Metadata Interchange between different systems created by OMG.

Because ATL supports M2M transformations between Ecore models, a requirement of having both models in Ecore format was made [10].

BPMN Ecore model was defined afterwards. In order to do that, a standard describing BPMN was analyzed and a set of usable elements was chosen. This set is a subset of all possible BPMN notation elements and an Ecore model was created for it.

Once two Ecore models were obtained, rules of transformation between CSV Ecore and BPMN Ecore models were defined using ATL tool. This was an important step of all conversion process confirming that formal table-based business process can be transformed to a BPMN. Afterwards, final step - transformation to a specific JBPM JPDL business process notation took place.

This transformation is required to make sure that the transformations are correct by comparing the result to other business processes used in author's workplace. Such business process would be used in further programming and implementation work as well. Afterwards, analysis of the resulting process was performed and an Ecore model was created. Transformation rules had to be defined as well. The final result of this process is an XMI document describing the JBPM JPDL business process of author's workplace. Graphical process of transformation is shown in Fig. 4.

A detailed overview on the process of transformation is shown next.

1. The sample business process is formed as a CSV document.
2. An ATL transformation is performed on it to get a document of XMI format corresponding to the CSV Ecore model.
3. Rules of checking CSV XMI document were formed by using oAW Check Constraints tool for Ecore models.
4. Using this tool, CSV XMI document is statically checked if the initial document has any errors.
5. Afterwards, the CSV XMI document is transformed to BPMN XMI document by using ATL transformation rules.

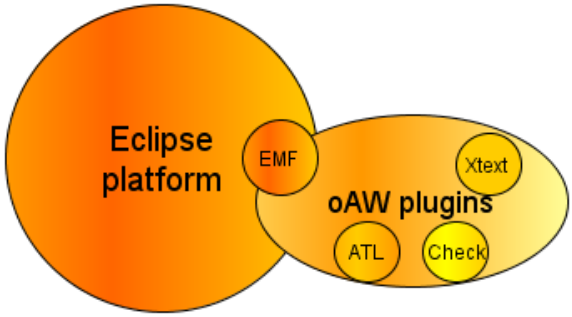

Fig. 3. Technological environment of prototype implementation.

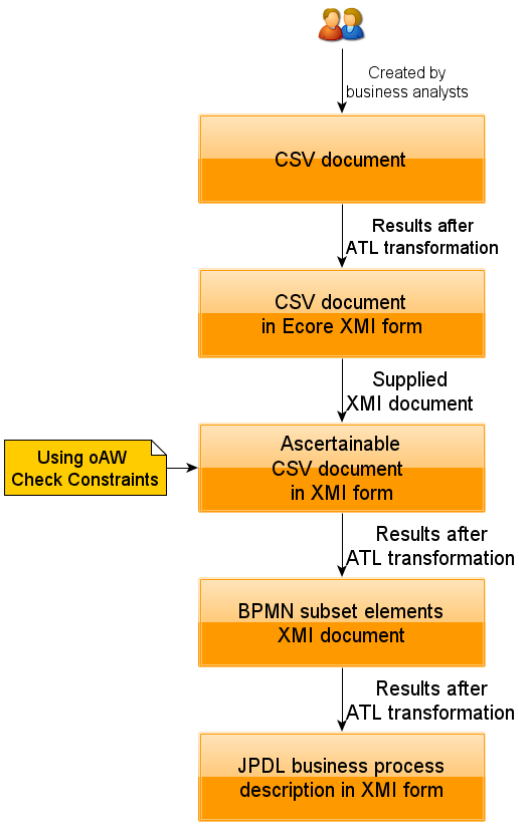

Fig. 4. General transformation process realized in prototype.

The final transformation of BPMN XMI document to JBPM JPDL business process notation is performed, resulting in a XMI document of specific business process notation. A prototype based on Eclipse plugins was realized. The Ecore models and transformation rules used will be presented.

2.2. Ecore Model of CSV Document

A structure of text-based business processes was specified in order to describe a meta-model of CSV document. Formal analyzing of table structure must be done. After analysis of a table structure used for business process notations, a new, more simple and more convenient structure was created, which is shown in Fig. 5.

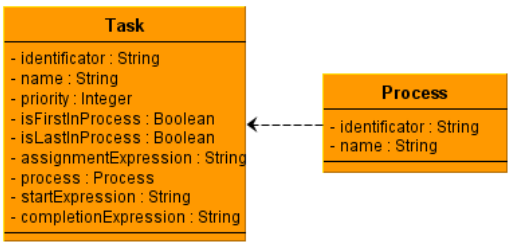

Fig. 5. Metamodel of table-based business process.

The description of fields of formal structure is shown in Table 1.

Rules of business process description table. The data presented in a table of the last chapter have to meet these conditions.

1. Task number field has to be unique for every process.
2. Task name has to be unique in a single process.
3. Task priority has to be between 1 and 5.
4. Every task except last has to have their initial or creation conditions.
5. Task end condition of the last task should not have any other tasks.

These are the basic and initial formal table-based business process rules. The advanced rules that need to be met by business process descriptions are presented below.

1. If a task is not the first one, its initial conditions have to be the same as the end condition of a previous task.
2. If there are several beginning tasks in a process, their initial conditions cannot overlap which means that several initial conditions cannot be met by the same set of data.
3. A process must have at least one first and last task.
4. The same task may not be used in several processes.
5. Initial (beginning) and final conditions may not overlap with one task.

As we can see, every rule is for initial or end conditions because the most of errors are found in these parts. With the help of successful implementation of verification of these parts, the time it takes to model these processes may be reduced significantly.

Table 1. Description of fields of the formal structure.

Table field	Description
Task number	Unique task number for every process.
Name or number of a process	Grouping of tasks to a process.
Task name	Unique task name in a process.
Priority	Task execution priority, a number between 1 and 5.
Initial/task creation conditions	Conditions describing the creation of a task. Comparison of initial attributes with fixed values in a condition.
Task completion conditions	Conditions upon which the current task should be finished and another task created. If there are no more tasks, the process itself is finished.
Is there another task in queue? (Y/N)	Marks whether this is the last task of a process.
Is this task the first one? (Y/N)	Marks the first task of a process.
Expression of task assignation to a user.	An existing username used to login to system is set as an expression value.

```
grammar org.xtext.example.\\
CsvDsl with org.eclipse.xtext.common.\\
Terminals generate csvDsl "CsvDsl"\\
Model :\\
(string+=string)*;\\
string :\\
          number_of_Task = ID';'\\
             processName = STRING';'\\
                taskName = STRING';'\\
     contitionOfCreation = STRING';'\\
ExpressionOfDefinitionTo = STRING';'\\
                 runTime = STRING';'\\
                priority = INT';'\\
finalysingConditions     = STRING';'\\
isTheLastTaskInProcess   = ('Yes'|'No');'\\
isTheFirstTaskInProcess  = ('Yes'|'No');'\\
```

Fig. 6. Example of the textual business process DSL.

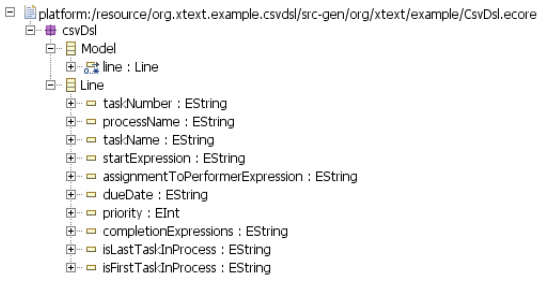

Fig. 7. Ecore model of text-based business process.

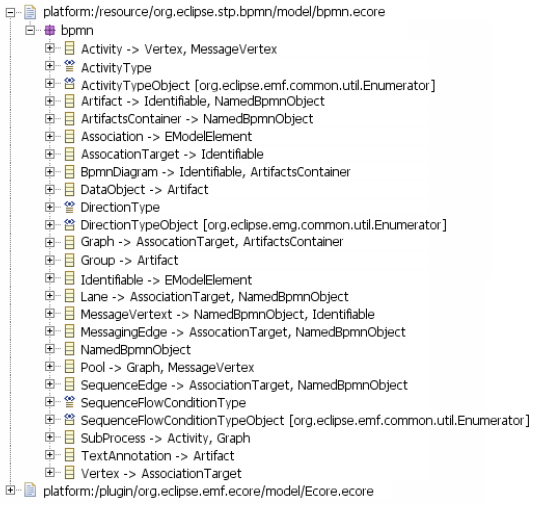

Fig. 8. Ecore model of BPMN elements.

Meta-model of CSV Document. DSL of a business process described using Xtext plugin is shown in Fig. 6. Even complex business processes may be described in such simple structure. Fig. 7 represents an Ecore model created from this DSL using Eclipse plugin. This CSV document model will be used in further M2M transformation process, where this model will be input (source), and standard description business process notation will be transformation output (objective) model.

3. Ecore Model of BPMN Elements Subset

Fig. 8 represents a set of elements and its Ecore model resulted after performing BPMN analysis. BPMN Ecore model was defined using BPMN standard [11]. This BPMN Ecore model is used for semantic BPMN. It is obvious that there should be another model involving depicting of elements (fonts, font sizes, element colours etc.) in business process diagrams, because BPMN is a graphical business process notation. A BPMN plugin using Eclipse platform, creates two documents at once, when creating BPMN business process diagram. One of them is semantic model, while the other is a document containing graphical element parameters. Since graphical positioning of elements is not important, it will be left aside and only semantic document and its structure will be analyzed.

3.1. JBPM JPDL Business Process Notation Ecore Model

The final model transformation result is a JBPM JPDL business process notation document, recognizable by Eclipse plugins. Fig. 9 represents Ecore model describing the document structure for this transformation.

By performing M2M transformation on BPMN Ecore model, a resulting JBPM JPDL Ecore model with XMI document format is received and used in further programming stage for final implementation to current system. Rules of transformation predicting details of the process have to be defined. These details consist of representation of an element from one model to another.

Fig. 9. Ecore model of JBPM JPDL notation elements.

3.2. Transformation Rules

The Ecore models described earlier do not perform any functions themselves, so transformation rules for transformation between models have to be defined. The following M2M transformations were described by using ATL tool and its Eclipse plugin:

i) from CSV Ecore to BPMN Ecore models;
ii) from BPMN Ecore to JBPM JPDL notation Ecore models;

ATL tool enables definition of transformation rules between a source Ecore model and a objective Ecore model bonding elements between them. An example transformation rule for converting a business process name from CSV Ecore to BPMN Ecore model is shown in Fig. 10.

```
rule ProcessName {\\
from\\
s : CsvDsl!Line\\
to\\
t : bpmn!BpmnDiagram (\\
title <- s.processName\\
```

Fig. 10. An example of a transformation rule.

The transformation rules were defined and a JBPM JPDL business process notation XMI document was received. This experiment shows that a business process described by a formal structure table may be converted to a business process notation of standard description type.

Results and Conclusions

Successful achievement of goals and tests of prototype leads to following statements:

i) formal structure tables may be used instead of standard business process notations to describe business processes;
ii) formal structure tables can be transformed to standard description business process notations;
iii) created prototype enables automation of conversion between formal table business process and a standard description business process.

The created solutions and achieved results may be applied in any business where business processes are not modeled using standard notations. By applying presented ideas and structuring initial business process tables to a structured row-column structure, transformations to a standard business process notation may be performed.

The described solution may be improved by expanding business process table structure, as well as increasing amount of fields moved from it to a middle BPMN notation, as well as defining transformation rules for several business process notations instead of one. That way, business processes could be transformed to more business process notations, as well as the sketches received after transformations would be fuller.

Acknowledgments

This work was prepared as a part of the Master Thesis. Author would like to thank Prof. Rimantas Vaicekauskas (Vilnius university) for valuable comments and suggestions.

References

1. Sigitas Vilkelis. Modeling of Business Processes. 1. Overview of Models and Metamodels – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 9-15.
2. Igor Sacevski, Jadranka Veseli. Introduction to Model Driven Architecture (MDA), Seminar Paper – 2007-06, 15p.
3. Wei Wang, Hongwei Ding, Jin Dong, and Changrui Ren. A Comparison of Business Process Modeling Methods – *IEEE* 2006, 6p.
4. Wang Hongxiu, Wang Gang, Wen Xiaoxian, Gao Guoan. Business Process Modeling for Multi-enterprise Cooperation – *IEEE* 2007, 4p.
5. Maryam Razavian, Ramtin Khosravi. Modeling Variability in Business Process Models Using UML – *IEEE* 2008, 6p.
6. A. Kleppe, J. Warmer, W. Bast. MIDA Explained. The Model Driven Architecture: Practice and Promise – Addison-Wesley, 2003, 170 p.
7. Dagmar Auer, Dirk Draheim, Verena Geist. Extending BPMN with Submit/Responsestyle User Interaction Modeling – IEEE Conference on Commerce and Enterprise Computing, 2009, 7p.
8. Eclipse Foundation. Model To Model – <http://www.eclipse.org/m2m/>, accessed 2010-05-01.
9. Eclipse Foundation. Eclipse Modeling Framework Project – <http://www.eclipse.org/modeling/emf/>, accessed 2010-05-01.
10. OMG. ATL – Atlas Transformation Language – <http://www.eclipse.org/m2m/atl/>, accessed 2010-05-01.
11. Object Management Group. Business Process Model and Notation (BPMN). Version 1.2, 2009-01-03 – <http://www.omg.org/spec/BPMN/1.2/PDF/>, accessed 2010-05-03.

CROWN: Applied tool for CARS-to-Raman spectrum decomposition

Ildar Galikov [a], Alytis Gruodis
[1] Faculty of Physics, Vilnius University, Saulėtekio 9, korp. 3, Vilnius, Lithuania

Received 2 Februar 2011, accepted 10 May 2011

Abstract. Novel advanced tool *CROWN* was created for CARS spectrum decomposition purposes: in order to extract Gauss- or Lorentz-shaped components containing Raman spectrum. *Win32* type application was created in object-oriented programming (OOP) manner using Visual Studio 8.0 package. C++ language was used for programming GDI interface using standard windows.h. library.

Citations: Ildar Galikov, Alytis Gruodis. *CROWN*: Applied Tool for CARS-to-Raman Spectrum Decomposition – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 22-26.

Keywords: *CROWN*; CARS; CARS-to-Raman; Spectrum decomposition.
PACS: 42.65.Dr
Short title: *CROWN*: CARS-to-Raman.

Introduction

Coherent anti-Stokes (AS) Raman scattering (CARS) technique [1] is popular as a unique macroscopic as well as microscopic [2] tool in material sciences [3], biology [4], biophysics [5], medicinal physics [6], etc. The benefit of CARS technique could be described as sensitivity to intramolecular changes and versality due to χ^3 behaviour [7]. CARS signal according to the experimental schema is blue-shifted from laser excitation frequencies – it means that CARS signal could be easily detected in the presence of strong luminescence [8].

Visualization of the digital data containing spectral dependencies belongs to the one of the most important computer tasks in computational physics. Information graphics software allows to users the fast manipulation in order to create the suitable visualization form from equipment-provided data-set. Computer-algebra systems (such as Maple [9] , MathCad [10] and Matematica[11]), numerical-software systems (LabView [12] , MatLAb [13] , SciLAb [14]) are quite useful for creating the two-dimensional or three-dimensional space projection containing object-oriented dependencies.

However, sometimes all functional possibilities of mentioned packages are not partially or fully required due to specificity of the task. Reordering and pre-manipulation (including sorting, fitting etc.) – such non-trivial operations are possible using Origin [15] package which, however, requires many non-automated operations. Using of Origin tools - worksheets and data fitting wizards – complicates the fast search of solution, but is very useful for slowly routine operations only.

This work is devoted to create the novel fast tool for express estimation of CARS and Raman spectrum content without necessity to use routine, time-wasting fitting procedure in the some wizards.

1. Physical behaviour of CARS spectrum

Project *CROWN* was created as original graphical user-friendly tool for fast Gauss/Lorenz-shaped decomposition of CARS spectra $S(\omega)$ into Raman components (CARS2Raman). Adapted formula from Ref. [16] was used as follows:

$$S(\omega_{AS}) = [I_{BCG}(\omega_{AS}) + I_R(\omega_{AS}) \cdot \cos\phi]^2 \qquad (1)$$

I_R and I_{BCG} represent the terms of Raman bands and backgrounds, respectively. Due to experimental conditions, such two terms represent different origin (changes of induced dipole moment for I_R(pure nuclear movement) and changes of nuclear dipole moment in the fast electrostatic field for I_{BCG} (nuclear movement in the electron environment), respectively). Factor $\cos(\phi)$, when ϕ=0 deg or ϕ=180 deg, represents the increasing or decreasing of intensity S. Angle ϕ represents the difference of phases between Raman term and background term.

Single Raman bands I_R could be described using classical Gauss I^G, Lorenz I^L or Voight I^V dependencies according to physical model of the task:

$$I^V = x \cdot I^G + (1 - x) \cdot I^L, \quad x \subset [0 \div 1] \qquad (2)$$

[a]Corresponding author, email: *Ildar.Galikov@gmail.com*

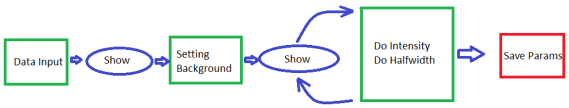

Fig. 1. *CROWN*. Setup of package.

For I^G and I^L dependencies, A_0 represents the altitude – intensity at ω_0, and σ represents the halbwidth of band.

$$I^G = \frac{A_0}{\sigma}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{2(\omega-\omega_0)^2}{\sigma^2}\right) \quad (3)$$

$$I^L = \frac{2A_0}{\pi}\frac{\sigma}{4(\omega-\omega_0)^2+\sigma^2} \quad (4)$$

2. Structuric scheme of *CROWN*

CROWN is written using *object-oriented programming* (OOP) language well known JAVA [17]. Java programming language was selected, due to it's portability between different operation systems. Different classes of program are responsible for data storage, calculating, visualization and *Graphical User Interface* (GUI). Therefore it creates the possibility for rapid updating of program.

Fig. 1 shows the principal scheme of the program. Firstly, input data as two-dimensional distribution is read and visualized. After that, user sets the background values. Next, user changes half-width or intensity of bands, in order to fit experimental and theoretically calculated from bands and half-width CARS spectra. When the purpose is achieved, bands parameters are writing to file. Used libraries and their description are shown in Table 1. The most important classes and sub-classes are shown and described in Table 2.

3. Description of graphical User Interface

Program *CROWN* is written to give a possibility to do easy, comfortable and fast CARS spectra decomposition in OPUS manner [18]. CARS spectrum decomposition proceeds from opening data file with experimentally calculated values, presented as X, Y matrix.

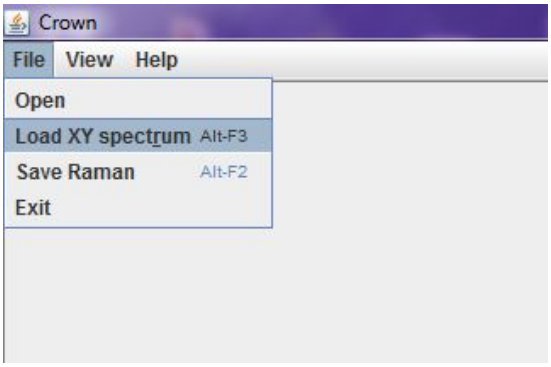

Fig. 2. *CROWN*: initial dialog for file input/output.

Fig. 2 shows file choosing dialog. The same result could be achieved by pressing "alt+F3" keyboard keys combination. After reading file, user must select background values, by pressing mouse at two points. The background is a linear function that goes through selected two points.

Table 1. List of used JAVA libraries.

GUI objects and their extended options
javax.swing.*
java.awt.*
javax.swing.border.*
javax.swing.JFileChooser
javax.swing.filechooser.FileNameExtensionFilter
javax.swing.event.DocumentListener
javax.swing.event.DocumentEvent
java.awt.event.*
File data management: Input/Output
java.io.BufferedInputStream.*
java.text.DecimalFormat
java.util.Scanner
java.io.File
java.io.FileInputStream
java.io.FileNotFoundException

Table 2. The most important classes and sub-classes.

Class name	Sub-class name	Description
WindowClass		GUI objects creation, response to user interactions
	MouseMoved	Mouse motion events
	MouseClicked	Mouse pressed events
	JInfoPanel	Output of important information to the screen(the right part of the workspace)
	PaintPanel	Spectrum data visualization, alongside other painting work.
	MenuLoadXY	Reading data from file.
	SaveAs	Writing data to file.
Bands		Data of a list of bands, for Raman and CARS spectrums calculation
Spectrum		Spectra (CARS, RAMAN, calculated CARS) data.
ChooseDialog		Dialog window, for changing bands data(x,y, phase).

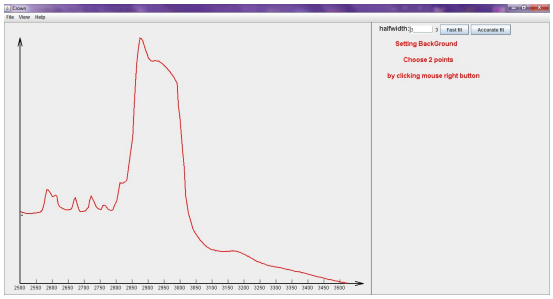

Fig. 3. Selecting of the first point of the background. Red curve corresponds the input file values.

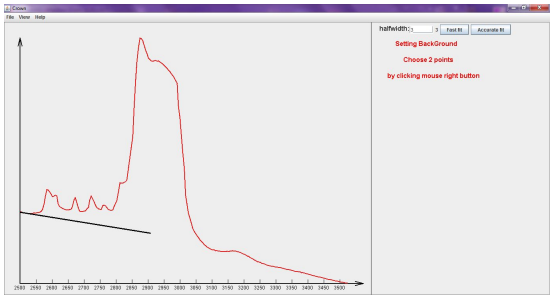

Fig. 4. Selecting the second point of background.

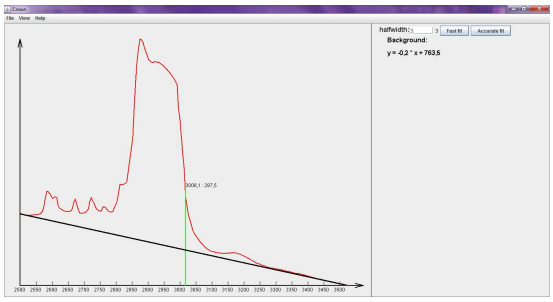

Fig. 5. After selecting both background points. At the right side of the workspace is written background equation. The black line corresponds to the background.

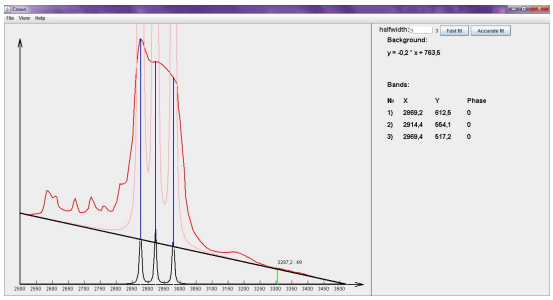

Fig. 6. Blue lines corresponds to the selected bands. Their values are shown in the information part of the workspace. Pink curve corresponds to theoretically calculated CARS spectrum. Black curve corresponds to the Raman spectrum.

Fig. 3÷10 illustrates visualization and fitting processes. Fig. 3, 4, 5 show the whole background defining process. Green line shows the place, where the band would be added, if the left mouse button be pressed. Fig. 6 shows experimental CARS spectra data, selected background and 3 bands. Bands can be deleted by pressing right mouse button. The closest to the mouse band (by x coordinate) will be deleted.

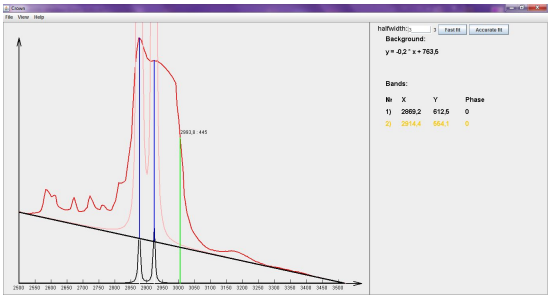

Fig. 7. Result after pressing right mouse button on the right side of all bonds.

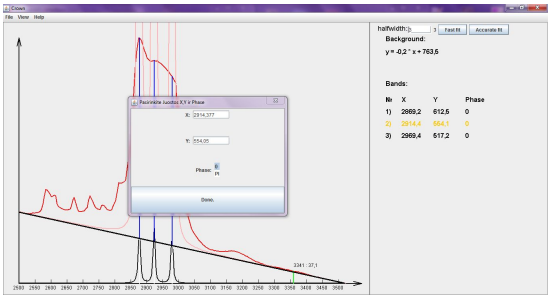

Fig. 8. The second band values changing. At the information part of the workspace , orange colour shows the band values are going to change.

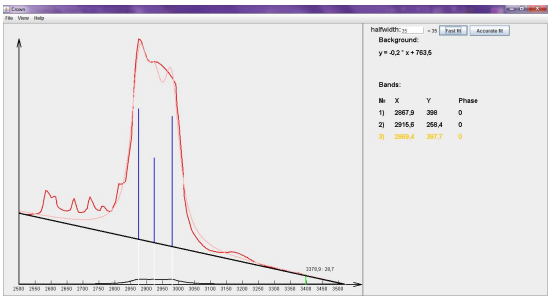

Fig. 9. results, after setting half-width = 35, and pressing “fast fit” two times.

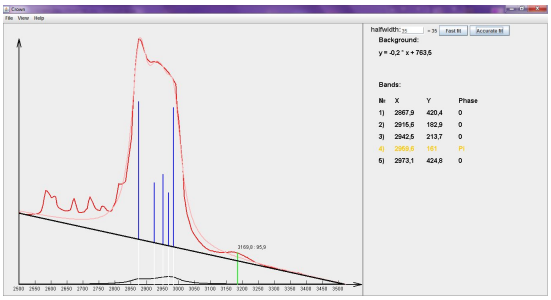

Fig. 10. Theoretical and experimentally calculated CARS spectrums after fitting.

Fig. 7 shows the result after pressing the right mouse button on the right side of all bands, presented Fig. 6. There is a possibility to change x, y and phase of any bond. To do that, user could either press middle mouse button(this selects the nearest band) or press the left button directly on the band (blue line). Fig. 8 shows the dialog window, of changing values of the middle band.

Band values could be fitted by pressing "fast fit", "accurate fit" buttons or changing halfwidth values. Fig. 9 shows the results, after setting half-width = 35 and pressing "fast fit" two times. In purpose to get less difference between experimentally and theoretically calculated CARS spectrums, use "accurate fit" button. Fig. 10 shows the better fitting, after adding two more bands and using "accurate fit" algorithm. Raman spectrum information could be exported by pressing "alt+F2" key combination, or by selecting the corresponding menu. Fig. 11 shows exported data. The first column shows X values, the second – Raman spectrum values. Other columns have values of Raman spectrum components each component place an accentual row in the Raman spectrum.

4. Fitting algorithm

It is important to understand, that CARS spectra changes, whenever band's phase or intensity are changed. The fitting is done, by changing Y values of bands, depending on difference between experimental and theoretically calculated CARS spectre values. It would take a huge amount of computer resources to recalculate CARS and Raman spectrums after changing any bands value. To solve that problem were created two fitting algorithms – fast fitting and accurate fitting. Algorithms start to work after pressing „do fit" or „accurate fit" buttons.

Fast fitting algorithm. This alhorithm is dedicated to fit approximately values. By running this algorithm every bands value changing by 5%, depending on difference of theoretically calculated value and value from inputted file at the bands X value. Theoretical CARS and RAMAN spectrums recalculated, after changing every bands value. This process is repeated thirty times.

Accurate fitting algorithm. This algorithm is dedicated to make theoretically calculated CARS spectrum as much closer to experimentally obtained spectrum value. It is obvious that bands with higher value have stronger influence on whole spectrum, so all bands are sorted by their value. CARS spectra recalculated right after changing any bands Y value. Therefore each other consequent calculations include corrective from all others. That is why this algorithm is much more accurate. Beginning from band with highest Y value and finishing with the lowest, band's value is changing by 1% and Raman and CARS spectrums are recalculated. This process is repeated ten times.

Conclusions

Novel tool was created for express estimation of CARS and Raman spectrum content without necessity to use routine, time-wasting fitting procedure in the some wizards.

Acknowledgements

This work was partly supported by the Lithuanian State Science and Studies Foundation (project B-07013 "KARS-KOPAS").

References

1. G. Knopp, Iddo Pinkas and Yehiam Prior. Two-dimensional time-delayed coherent anti-Stokes Raman spectroscopy and wavepacket dynamics of high ground-state vibrations. – *Journal of Raman spectroscopy* 31 (2000) 51–58.
2. Mamoru Hashimoto, Tsutomu Araki, and Satoshi Kawata Multi-focus coherent anti-Stokes Raman scattering microscopy. – *Microsc. Microanal.* 9(Suppl 2) (2003) 1090-1091.
3. Josefa R Baena1 and Bernhard Lendl. Raman spectroscopy in chemical bioanalysis. – *Current Opinion in Chemical Biology* 8 (2004) 534–539.
4. M. O. Scully, G. W. Kattawar, R. P. Lucht, T. Opatrny, H. Pilloff, A. Rebane, A. V. Sokolov, and M. S. Zubairy. Fast CARS: Engineering a laser spectroscopic technique for rapid identification of bacterial spores. – *PNAS* Vol. 99 N.17 (2002) 10994–11001.
5. Haifeng Wang, Yan Fu, Phyllis Zickmund, Riyi Shi, and Ji-Xin Cheng. Coherent Anti-Stokes Raman Scattering Imaging of Axonal Myelin in Live Spinal Tissues. – *Biophysical Journal* Vol.89 (2005) 581–591.
6. Hilde A. Rinia, Mischa Bonn, Erik M. Vartiainen, Chris B. Schaffer, and Michiel Müller. Spectroscopic analysis of the oxygenation state of hemoglobin using coherent anti-Stokes Raman scattering. – *Journal of Biomedical Optics* Vol.11-5 (2006) 050502-1,3.
7. Schluker S., Schaeberle M.D., Huffman S.W., Levin I.W. Raman microspectroscopy: a comparison of point, line, and widefield imaging methodologies. – *Anal Chem* 75 (2003) 4312-4318.
8. Arenas JF, Lopez-Tocon I, Centeno SP, Soto J, Otero JC: How a resonant charge transfer mechanism determines the relative intensities in the SERS spectra of 4-methylpyridine. – *Vib Spectrosc* 29 (2002) 147-154.
9. Maple13, math and engineering software, <http://www.maplesoft.com/>, accessed 2009 11 15.
10. MathCad, engineering calculation software, <http://www.ptc.com/products/mathcad/], accessed 2009 11 15.
11. Matematica 7.0.1, computational software program for technical computing, <http://www.wolfram.com/products/mathematica/ latestversion/>, accessed 2009 11 15.
12. LabView, graphical programming for measurement and automation, <http://www.ni.com/labview/>, accessed 2009 11 15.
13. MatLAb, interactive environment to perform computationally intensive tasks, <http://www.mathworks.com/products/matlab/>, accessed 2009 11 15.
14. SciLAb 5.1.1, open source platform for numerical computation, <http://www.scilab.org/>, accessed 2009 11 15.

15. Origin 8.1, analysis tools for Spectroscopy, Statistics, Signal Processing, 3D Surface Fitting and Image Processing, <http://www.originlab.com/>, accessed 2009 11 15.
16. Erik M. Vartiainen, Hilde A. Rinia, Michiel Müller, Mischa Bonn. Direct extraction of Raman line-shapes from congested CARS spectra. – *Optics express* 14(8) (2006) 3622.
17. Java SE Development Kit 6 (JDK 6) 25 update.
18. OPUS viever 6.5, Bruker Optics spectroscopic software, <http://www.brukeroptics.com>, accessed 2009 11 15.

Zipf and Related Scaling Laws. 1. Literature Overview of Applications in Economics

Artūras Einikis [a] , Giedrė Būdienė, Alytis Gruodis
Vilnius Business College, Kalvarijų str. 125, Vilnius, Lithuania

Received 31 Januar 2011, accepted 12 March 2011

Abstract. Zipf law is well-known by modeling the economic human activity. Power law distributions of such type (so called zipfian) through parametrization are related to the Pareto distribution. This review is devoted to analysis of application of scaling law in economics. The six themes are observed: company size and bankruptcy, wealth distribution, systems of finite recourses, investment strategy, trading and stock market models, city creation mechanism and driving forces for city expansion.

Citations: Artūras Einikis, Giedrė Būdienė, Alytis Gruodis. Zipf and Related Scaling Laws. 1. Literature Overview of Applications in Economics – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(11)** 2011 – Pp. 27-35.

Keywords: Zipf law; power law; long tail; scaling.
Short title: ZIPF law - economics - 1.

Introduction

Power-law dependencies are very large distributed in the world. Many sets of data studied in economical and natural sciences can be approximated by the dependence where probability is inversially proportional to the item rank. Many factors in economics such as executive pay, income, trading volume, international trade, wealth, stock market returns, the size of cities and firms etc are surprisingly distributed according to power law with exponent equal to 1.

As an example, power law is well known in linguistics. Items of large regular texts written in any human language are distributed not randomly but follow a power law.

George Zipf [1] found that frequency $f(r)$ of item (word) occurrence in finite corpus is inversely and linearly related to item rank $r(w)$ – so called **Zipf** law, see Eq.(1).

$$f(r) = \frac{\alpha}{r^{\gamma}} \tag{1}$$

Benoit Mandelbrot [2] proposed the generalized expression of Zipf distribution as a discrete probability distribution – so called **Zipf-Mandelbrot** distribution.

$$f(r) = \frac{\alpha}{(1+\beta r)^{\gamma}} \tag{2}$$

$$f(r) = \frac{\alpha}{\beta + r^{\gamma}} \tag{3}$$

Both distributions contain the adjustable parameters α, β, γ which are item content-dependent. In particular case for ranked frequency $f(r)$ of item occurrence in finite English corpus, Zipf distribution parameters are following: $\alpha \approx 0.1$, $\gamma \approx 1$ (Eq.(1)). Fig.1 represents idealized single-linear Zipf dependence in log-log scale.

In theory of statistics, the **Yule–Simon** distribution is a discrete probability distribution.

$$f(r) = \frac{\alpha \cdot \beta^{r}}{r^{\gamma}} \tag{4}$$

Eq.(4) represents the limiting distribution of a particular stochastic processes which was studied by Udny Yule as a distribution of biological objects. Herbert A. Simon [3] rationalized mentioned compound distribution where the parameter of a geometric distribution was treated as a exponential function.

Exponential function – see Eq.(5) – is useful for comparative analysis when function growth (decay) rate is proportional to the function argument. In many cases exponential model could be treated as the fundamental due to relations to the Gauss normal distribution.

$$f(r) = \frac{\alpha}{\exp(\gamma r)} \tag{5}$$

For multi-level complex system, second generalization of Zipf law was realized as well-known **Menzerath-Altmann** equation [4] – see Eq.(6).

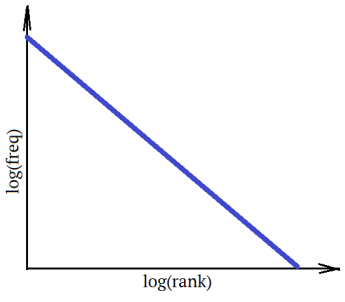

Fig. 1. Frequency dependence on rank. Idealized single-linear Zipf distribution in log-log scale - see Eq.(1), γ=const.

[a]Corresponding author, email: *arturas@kolegija.lt*

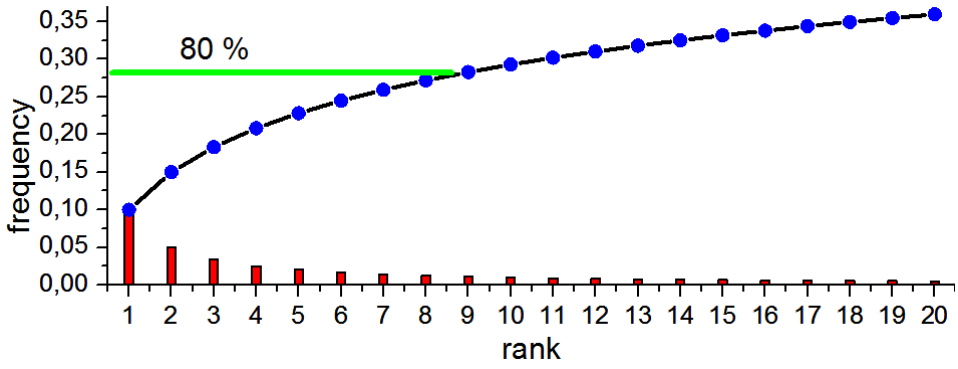

Fig. 2. Frequency dependence on rank. Idealized single-linear Zipf distribution, Eq. (1), α=1, γ=1. column bar, red. Pareto chart, blue. 20% of products gives 80% of profits (Pareto rule).

$$f(r) = \frac{\alpha \cdot r^{\beta}}{\exp(\gamma r)} \tag{6}$$

Gibrat law claims that size of firm and its growth rate are independent [5]. Vilfredo Pareto observed in 1906 that 80% of the land in Italy was owned by 20% of the population. The **Pareto** principle (also known as the principle of factor sparsity, so-called factor scattering or 80/20 rule) states that, for many events, roughly 80% of the effects come from 20% of the causes.

$$f(r) = \left[\frac{r}{r_{min}}\right]^{-\gamma} \tag{7}$$

Fig. 2 represents Zipf distribution (column bars represent the ranked frequencies - individual values in descending order) and Pareto chart (line graph represents cumulative total index). The purpose of Pareto chart is to highlight the most important amount of ranked factors. In many cases, 80% of content could be titled as significant. Pareto chart belongs to the famous tools of quality control.

Gutenberg–Richter law [6] expresses the relationship between the magnitude M and total number of earthquakes N in any given region and time period.

$$\log_{10} N = \alpha - \beta \cdot M \tag{8}$$

Eq.(8) was derived originally in seismology from empirical data. Modern attempts in explanation are grounded on self-organized criticality.

Hill estimator [7] is a popular method for estimating the thickness of heavy tails. Approximation of the distributional tail must be provided with a power function. In practice it is often true equation Eq.(9) for x>0:

$$P(X > x) \approx Cx^{-\gamma} \tag{9}$$

Then the idea is to estimate the parameters $C > 0$ and $\gamma > 0$ by a conditional maximum likelihood estimate based on the r+1 (0<r<N) largest order statistics, which represent only the portion of the tail for which the power law approximation holds.

Usage of Hill estimator in some cases is sophisticated but sometimes it is necessary due to so called robustness of dependencies. Since it only depends on the shape of the probability tails, it can be applied in situations where the form of the distribution is unknown. This is typically the case in applications to finance, where heavy tails are common.

Zipf law in economics is well-known by modeling the ranked firm size distribution, income-wealth distribution, city size distribution etc. Power law distribution of such type (so called zipfian) through parametrization are related to the Pareto distribution. This review is devoted to analysis of application of scaling law in economics. Six themes of big importance are observed here:

1) company size and bankruptcy;
2) wealth distribution;
3) resourses and investment strategy;
4) trading / stock market models;
5) city creation mechanism;
6) driving forces for city expanding.

1. Company size and bankruptcy

Economic prosperity is determined by the activity of the firms. Firms, stock companies, corporates etc are established to achieve certain economic goals for a certain period of time. Creation, growth, prosperity, stagnation, and bankruptcy - these states of company describes the natural way of development, which influences the macro-economic indicators. Although firm growth and bankruptcy are stochastic processes, they could be forecasted by analysing dynamical tendencies in certain economic area.

Ausloos et al. [8] state that many problems in economy and finance is possible to solve using methods of statistical physicists. Presence of financial cycles and existence of power-law correlations in economic systems allow to use digitalized methods such as fluctuation analysis, multi-component analysis etc. The well-known financial analyst technique, moving average, is shown to raise questions about fractional brownian motion properties. Also Zipf method is useful for sorting out

short range correlations.

Wright [9] represent the self-organized dynamic model of the social relations between workers and capitalists. Several empirical distributions were used: power-law firm size distribution, the Laplace firm and GDP growth distribution, the lognormal firm demises distribution, the exponential recession duration distribution, the lognormal–Pareto income distribution, and the gamma-like firm rate-of-profit distribution. In the framework of model, these distributions are interconnected in order to generate the business cycle phenomena. The generation of an approximately lognormal–Pareto income distribution and an exponential–Pareto wealth distribution demonstrates that the power-law regime of the income distribution can be explained by an additive process on a power-law network that models the social relation between employers and employees organised in firms, rather than a multiplicative process that models returns to investment in financial markets.

Europe and USA. Firm growth could be modeled as clustering process. Clustering of large number objects was provided by means of Zipf and Yule distributions [10]. Gibrat rule of proportionate growth claims that size of firm and its growth rate are independent. In many cases, Girbat law contains empiric error due to stochastic grown process [11].

Galeo et al. [12] analyse very large amount of data containing G7 group's firms over the period 1987÷2000 in several business cycle phases. Power law distributions are satisfied in all cases, but differences between parameters related to the recession and expansion processes are significant (the exponent $\gamma \rightarrow 1$, i.e., the resulting size distribution generally is not zipfian).

Axtell [13] analyses the distribution of USA firm sizes at historical perspective. Zipf distribution at lognormal scale takes place. Wyart et al. [14] studied Sutton 'microcanonical' model for the internal organization of firms as an alternative model based on power-law distribution. In that case, growth rates are asymptotically gaussian, whereas empirical results suggest that the kurtosis of the distribution increases with size.

Amaral et al. [15] analyse the Compustat data base comprising all publicly-traded United States manufacturing firms within the years 1974-1993. Amaral concludes the distribution of the logarithm of the growth rates, for a fixed growth period of one year, and for companies with approximately the same size, displays an exponential form.

Asia. Taking into account the parameter expressing the size of firm, power law expressions allow to receive the correlated distributions involving both processes: destruction and creation [16-17]. Some kinematical relationships between Pareto–Zipf and Gibrat laws are presented by Fujiwara [18]. Fujiwara et al. [16] analyse large number of European firms using power-law dependencies. Upper-tail of the distribution of firm size can be fitted with Zipf dependence, and that in this region the growth rate of each firm is independent of the firm's size. This sentence satisfied the Gibrat law.

Zhang et al. [19] analyse the data of top 500 Chinese firms from the year 2002 to 2007. Dependence of firm size on rank is presented according to Zipf law (exponent γ=1 for each year). Phenomenon explanation of it based on a simple economic model which takes capital accumulation into account.

Gupta et al. [20] studied the statistical distribution of firm size for USA publicly traded firms through the Zipf plot technique. Sale size is used to measure firm size. The log-normal distribution has to be gradually truncated after a certain critical value for USA firms. Therefore, the original hypothesis of proportional effect proposed by Gibrat is valid with some modification for very large firms.

Bankruptcy. Byoung Hee Hong et al. [21] studied the scaling behaviours for fluctuations of the number of Korean firms bankrupted in 2002-2003. Power law distribution of the number of the bankrupted firms takes place and Pareto exponent is close to unity.

Fujiwara [17] studied the data of Japanese bankruptcy in 1997. Zipf law dependencies could be estimated for the distribution of total liabilities of bankrupted firms in high debt range. The life-time of these bankrupted firms has exponential distribution in correlation with entry rate of new firms. Debt and size are highly correlated, so the Zipf law holds consistently with that for size distribution.

2. Wealth distribution

Souma [22] reported empirical studies on the personal income distribution, where two models were used: lognormal and power law. Pareto and Gibrat indexes were used as an unversal factors in order to estimate the temporal changes.

Europe and USA. Pareto distribution was devoted to describe the allocation of the wealth among individuals. In any society at any times the larger portion of the wealth (80% by Pareto [23], 70% by Gide [24]) is owned by a smaller percentage of the people (20% Pareto, 30% by Gide).

Hegyi [25] analyses the distribution of wealth in the medieval Hungarian aristocratic society. Wealth distribution was find according to power-law nature. Using no-trade limit of wealth-distribution model, Pareto law validity was confirmed for feudal society. Obtained Pareto exponent $\gamma \in [0.92 \div 0.95]$ is closed to 1.

Iglesias et al. [26] analyse the emergence of Pareto wealth power-law distribution. Models including the risk factor were proposed and tested. For constant risk aversion the system self-organizes in a distribution that goes to a Gaussian. Surprisingly, it was established that random risk aversion can produce distributions going from exponential to log-normal and power-law. Correlations between wealth and risk aversion was found.

Parameterization using temperature model occurs by solving unregular tasks of large scale wealth distribution. Dragulescu et al. [27] analyse the data on wealth and income

distributions in the United Kingdom, as well as in several states of the USA. Great majority of population is described by an exponential distribution, and the high-end tail follows a power law. New empirical parameter – temperature (as analogy in physics) was introduced in order to characterize "the kinetic energy" of society.

Hernandez-Perez et al. [28] analyse company size distribution for developing countries using the framework proposed by Ramsden et al. [29]. Not adequate living conditions not allow to compare the usual makroeconomic parameters such as Zipf exponent etc. Hypothesis of additional parameter which plays a role analogous to the temperature of the economy occurs after decision that the level of economic development must be estimated in usual power-law dependencies.

Ausubel [30] analyses the myth *Living like America* from the economic perspectives. He claims from historical point of view that incomes vary for the very simple reason: income crowns the successful completion of a series of multiplicative tasks, causing a skewed distribution. As incomes rise, however, economic, social, and environmental requirements and capacities grow.

Trigaux [31] describes the main principles for tasks of econophysics and economy simulations. Ground idea could be formulated as follow: the repartition of wealth in every economic system always occur in Pareto law. As this inegalitarian repartition is a cause of many problems in the world, it would be interesting to find a remedy. Econophysics studies always start from the hypothesis as what economy systems are formed only of agents perfectly egocentric, each seeking only to gather the maximum of wealth for himself. Trigaux formulated two questions:

i) should the Pareto law come only of this limiting hypothesis (the maximum of wealth for himself)?
ii) should the Pareto law come in case if agents had other types of behaviours, for instance altruistic?

In order to simulate the proposed situation, two behaviours (altruism and egocentrism) were parameterized. Really much more egalitarian repartition appears, even with a relatively low rate of altruism (15%). More so, this egalitarian repartition occurs according to a Gauss law which is completely different law from that of Pareto.

Asia. Okuyama et al. [32] analysed the distribution functions of annual income of companies. Power-law distribution (according to Zipf law) was confirmed. Aoyama [33] analysed personal income, company's income, and various measures of company size. Some relationships under the Pareto–Zipf law and Gibrat law of detailed balance were established as a basis for perturbative treatment of the economic change.

Isikawa et al. [34] analyse the database of high income companies in Japan. Quantitative relation between the average capital of the companies and the Pareto index was find. Quantitative relation between the lower bound of capital and the typical scale at which Pareto law breaks was established.

Theoretical study of the changes in poverty with respect to the 'global' mean and variance of the income distribution using Indian survey data was done by Chattopadhyay et al. [35]. Authors claim that Pareto poverty function satisfies all standard axioms of a poverty index presented by Kakwani [36] and Sen [37]:

i) monotonicity axiom: given other things, a reduction in income of a person below the poverty line must increase the poverty measure;
ii) transfer axiom: given other things, a pure transfer of income from a person below the poverty line to anyone who is richer must increase the poverty measure.

Evolutionary games represent an important factor in simulating of economic environment. Mao-Bin Hu et al. [38] proposed full-time study of wealth distribution with agents playing evolutionary games on a scale-free social network. Pareto power-law distribution is satisfied for agent's personal wealth prediction. Phenomenon of accumulated advantage (so called **Matthew** effect, the rich get richer and the poor get poorer) was validated also by analysing the agent's personal wealth correlation to its number of contacts (connectivity).

3. Resourses and investment strategy

Naldi [39] studied the relationships between Zipf law and the major concentration indices. Standard model where the firms' size are related to the financial investment amounts was used. It was established that Hirschman–Herfindahl index [40] is the most sensitive index in contexts where Zipf law applies. Applications of Zipf law could play an estimating role many very sensitive marked indicators.

Ausloos et al. [41] describe strategy how to apply the Zipf method to extract the γ-exponent for seven financial indices (DAX, FTSE, DJIA, NASDAQ, S&P500, Hang-Seng and Nikkei 225). Ausloos et al. [42] studied short-range time correlations in financial signals by means of Zipf method and the i-variability diagrams (VD). A precise Zipf diagram analysis has been shown to lead to a non-immediate information on the signal behaviour, even taking into account error bars.

Alegria et al. [43] probed to relate the parameters of Pareto-type distribution of bank sizes to the specific bank indexes such as Herfindahl–Hirschman index and the top 5%-concentration ratio. Effect of changes in Zipf exponent γ correlates to sample size. Wilhelm et al. [44] analyse an elementary stochastic model representing the system with finite resources where power-laws distribution takes place. This model extends the scale-free network model (SF) to include the fact of finite resources.

Saif et al. [45] investigate the problem of wealth distribution from the viewpoint of asset exchange. The simple asset exchange models (grounded on Pareto law) fail to reproduce trading strategies. Two models were used for successful simulation of trade:

i) Yardsale (YS) purpose model; and
ii) theft and fraud (TF) model.

Power-law tail in wealth distribution was observed in case if the agents are allowing to follow either of the mentioned models with some probability.

4. Trading / stock market models

The most important task is to create the dynamic market model which could predict the trade type and day-to-day fluctuations. Balakrishnan et al. [46] studied and modeled the distribution of daily stock trading using the power law. New phenomenon was established that the trading is becoming increasingly concentrated in a subset of stocks. The power law exponent systematically increases with time suggesting. Tuncay et al. [47] analyse the daily financial volume of transaction on the New York Stock Exchange and its day-to-day fluctuations. Gaussian distribution for longer time intervals, like months instead of days takes place. Otherwise, power-law tails could be attracted to long-term trends. Unconditional volatility distribution [48] of the Italian futures market were studied by Reno et al. Transactions in period of 2000 and 2001 (including event of dramatic 11 September 2001) were characterized by unusually high volatility levels. Results show that the standard assumption of lognormal unconditional volatility has to be rejected for such a turbulent sample, since it is unable to capture the tail behaviour of the distribution.

Gabaix et al. [49] presented a theory of excess stock market volatility. Market movements are due to trades by very large institutional investors in relatively illiquid markets. Power law distribution can be presented for resuming evaluation of trade, but optimal trading behaviours are stochastically-dependent.

Ideal-gas-like-models. Chatterjee et al. [50] reviewed big number of market models differing by shape of the distribution of wealth. Several paradigms from physics such as ideal-gas-like models of markets are observed across varied economies. Presented realistic model where the saving factor can vary over time (annealed savings) is yielding the Pareto distribution of wealth in certain cases. Numerical simulation presented in Ref. [51] describes the ideal-gas model of trading markets, where each agent is identified with a gas molecule and each trading as an elastic or money-conserving two-body collision. Unlike in the ideal gas, quenching/ saving properties are included. Model is showing self-organized criticality, and combines two distributions: Gibbs and Pareto.

Bhattacharyya et al. [52] obtained common mode of origin for the power laws:

i) the Pareto law was used for the distribution of money among the agents with random-saving propensities in an ideal gas-like market model; and
ii) the Gutenberg–Richter law for the distribution of overlaps in a fractal-overlap model for earthquakes.

Lotka–Volterra formalism. Market stability was studied using the generalized Lotka–Volterra (LV) formalism by Louzoun et al. [53]. LV equations are non-linear differential equations, pair of first-order, frequently used to describe the time-dependent dynamics of biological systems in which two species interact, one as a predator (y) and the other as prey (x):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - \beta xy \tag{10}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \delta xy - \gamma y \tag{11}$$

Parameters α, β, γ and δ describe the interaction of the two species. First derivatives of x and y represent the growth/decreasing rates of the mentioned populations over time.

Power law distributions in the individual wealth (according to Pareto law) and financial markets returns (fluctuations) show auto-catalytic or multiplicative random character of the capital dynamics. Exponent of the power laws turns out to be independent on the time variations of the average. This explains also the stability over the past century of experimentally measured Pareto exponent. Strong feedback signalizes the danger of the market stability.

Solomon et al. [54] adapted generalized LV model with mutiagent systems in order to investigate economic systems. Weak generic assumptions on capital dynamics were realized in model of predictions for the distribution of social wealth. In 'fair' market, the wealth distribution among individual investors fulfils a power law.

Simulations and games. Chebotarev [55] propose the study of a hierarchical income model for asymmetrical transactions: directions of money movement and commodity movement are opposite. The price-invariance of transactions means that the probability of a pairwise interaction is a function of the ratio of incomes, which is independent of the price scale or absolute income level. The income distribution is a well-defined double-Pareto function, which possesses Pareto tails for the upper and lower incomes. The Pareto exponents are also stable with respect to the choice of a demand function within two classes of status-dependent behaviour of agents.

Mohanty [56] presented an economy model by taking N independent agents who gain from the market with a rate which depends on their current gain. Power-law distributions take place. Kuscsik et al. [57] studied the model of environmental–economic interactions. The interacting heterogeneous agents are simulated on the platform of the emission dynamics of cellular automaton. Steady-state and non-equilibrium properties were established in such type simulation. Relationship to Zipf law and models of self-organized criticality were discussed.

Yanagita et al. [58] studied a simple model of market share dynamics with rational consumers and firms interacting with each other. Simulation results show that three phases of market structure appear depending upon how rational consumers

are. Three phases could be titled as the uniform share phase, the oligopolistic phase, and the monopolistic phase.

In an oligopolistic phase, the market share distribution of firms follows Zipf law and the growth-rate distribution of firms follows Gibrat law. An oligopolistic phase is the best state of market in terms of consumers' utility but brings the minimum profit to the firms because of severe competition based on the moderate rationality of consumers.

5. City creation mechanism

City growth phenomenon is well known from the Antic time as the parameter of civilization development. Growth is stimulated by the human activity in case if the resources are in enough amounts. City growth process could be described according to the power-law dependence as was checked in the middle of XX century. Formulated as an universally law, city size distribution was related to the power function. So called **Zipf law for cities** (exponent γ=1) was treated as the quit enough power law realization.

$$f(r) = \frac{\alpha}{r^1} \tag{12}$$

Nitsch [59] provides very large study of the empirical literature on Zipf law for cities including 515 estimates from 29 studies. Surprisingly, Zipf exponents are significantly larger than 1.0. This finding implies that cities are on average more evenly distributed than suggested by Zipf law.

Marsili et al. [60] presented a general approach to explain the Zipf law of city distribution. Benguigui et al. [61] presented an application of a growth model for a system of cities (computer model simulation). Model includes a random multiplicative process for the growth of individual entities and for the creation of new ones. Expression with a positive exponent -"shape exponent" and additional three parameters was used in order to describe the dynamics of the systems' size distributions through time. Quit good agreement at the macro level between the model and the real data takes place.

Pareto distribution allows to make the very strong city size estimation in many countries. Soo [62] solved the task of empirical validity of Zipf law for cities, using data on 73 countries. Two estimation methods - OLS (ordinary least squares) and the Hill estimator – were used. The OLS estimates of the Pareto exponent are roughly normally distributed, but those of the Hill estimator are bimodal. Variations in the value of the Pareto exponent are better explained by political economy variables than by economic geography variables. Cordoba [63] derived the conditions in the framework of Pareto model. Presented rules must satisfy the standard urban model:

i) a balanced growth path; and
ii) a Pareto distribution for the underlying source of randomness.

Gabaix [64] presented review surveys of well-documented empirical power law regarding income, wealth, the size of cities etc. Random growth, condition optimization must be treated as the adjustable parameters. Some empirical regularities currently lack an appropriate explanation. Gabaix also describes the open areas for future research.

City size represent a geometrical distribution of urbanized areas. Benguigui et al. [65] presented the growth model for a system of cities which is grounded by not only Zipf law but also other kinds of city size distributions. Power-law like function with exponent γ (for Zipf law γ=1) was introduced. Three classes of city size distributions depending on the value of γ were defined: i) γ>1; ii) γ<1; iii) γ=1. The model is based on a random growth of the city population together with the variation of the number of cities in the system. It was concluded that the exponent γ may be larger, smaller or equal to 1, just like in real systems of cities, depending on the rate of creation of new cities and the time elapsed during the growth. It is necessary to point out that the influence of the time on the type of the geometric distribution must be treated as significance.

Carvalho et al. [66] studied the distribution of the length of open space linear segments, derived from maps of 36 cities in 14 different countries. By scaling the Zipf plot of l, two master curves for a sample of cities, which are not a function of city size, were obtained. It means that third class of cities is obtained, and this class is out of classification order. According to Zipf plot, this distribution is realized in region of power-law tails with exponent γ=2. Small correlation between real data and the possibility of observing and modeling urban geometric structures was suggested. Volchenkov et al. [67] studied the distribution of open space in city. The area of open space which are related to the other spaces is distributed according to the power-law statistic. Observed universality may help to establish the international definition of a city as a specific land use pattern.

Stochastic model of city growth represent a behaviour of cluster formation type where time-dependent processes occurs. Zanette et al. [68] proposed stochastic model for govern city formation. The model predicts a power-law population distribution whose exponent is in excellent agreement with the universal exponent observed in real human demography. Zanette suggested that urban development at large scales could be driven by intermittency processes. Duranton [69] presented canonical model of endogenous growth with product proliferation into a simple urban framework (which yields Zipf law for cities). The stochastic outcomes of purposeful innovation and local spillovers can thus serve as foundations for random growth models.

6. Driving forces for city expanding

Mansury et al. [70] presented a spatial agent-based model to generate a system of cities that exhibits the statistical properties of the Zipf Law. Two main factors could be estimated as of most important significance: bounded rationality and

maximum heterogeneity of agents. Combination of such two factors can produce a generic power law relationship in the size distribution of cities, but does not always generate the dependency according to Zipf law. Zipf law breaks down unless the extent of agglomeration economies overwhelms the negative disagglomerating forces. Decker [71] probed to solve the city growth task when largest cities comprise the long tail of the distribution. In order to explore generating processes, simple model was used. Model incorporates only two basic human dynamics: migration and reproduction.

Semboloni et al. [72] presented the model for the distribution of individuals in cities. The number of individuals is fixed and the dynamic depends on migration from one city to another. Two strategies were used for modeling purposes: utilisation of resources for production and selling of products to people. The most important statements can be formulated as follows.

1. Because resources are uniformly distributed and shared among individuals, the first strategy pushes individuals in small cities - *unification*.
2. In turn, because selling depends on the quantity of individuals are living in a city, the second strategy pushes individuals in big cities - *diversification*.

Random application of unification and diversification strategies results in power-law distribution of cities.

Europe. Sarabia et al. [73] introduced the Pareto-positive stable distribution as a new model for describing city size data in a country. The mentioned distribution provides a flexible model for fitting the entire range of a set of city size data. The classical Pareto and Zipf distributions are included as a particular case. City size data for Spain for several different years was considered. The new distribution is compared with three classical models: Pareto, lognormal and Tsallis distributions.

Asia. Anderson et al. [74] analyse city size distribution in China using two behaviours: i) the relative growth of cities and ii) the nature of the city size distribution. This analysis was provided in the framework of political conditions such as Economic Reforms and the One Child Policy since 1979. It was established as a reason for the significant structural changes in the Chinese urban system. The city size distribution remains stable before the reforms but exhibits a convergent growth pattern in the post-reform period. It was concluded that log-normal rather than Pareto specification turns out to be the preferred distribution.

Gangopadhyay et al. [75] studied the size distributions of urban agglomerations for India and China. Authors have estimated the scaling exponent for Zipf law with the Indian data (1981-2001) and Chinese data (1990-2000). Parameters of Pareto and Tsallis q-exponential distribution have been estimated: for India, $\gamma \in [1.88 \div 2.06]$ and for China, $\gamma \in [1.82 \div 2.29]$.

Chen [76] examined the relation between the feature of increasing returns in the dynamic growth process and the property of power law in the static limiting distribution. Fractal-like structures used in this model implies both the power law and rank size rule. Power law or Zipf law are valid for the distributions of city size. Gibrat law proposes general and neat interpretations for this regularity in a city distribution, but the homogeneity assumption in Gibrat law shows a disregard of the agglomeration effect that is essential in economic interpretation. Path-dependent nonlinear Polya processes were appended to analyse the relation between the feature of agglomeration in the path-dependent processes and rank-size relations in the limiting distributions. Author conclude that the assumption of agglomeration economies must be significant. It allows to state that the agglomeration benefits increase without a ceiling as the residents are added to the city.

South America. Moura et al. [77] studied the application of Zipf law for cities distribution in Brazil. The results show that the population distribution in Brazilian cities does follow a power-law similar to the ones found in other countries. Values of the power-law exponent were found to be about $[2.2 \div 2.3]$. More accurate results were obtained with the maximum likelihood estimator, showing an exponent equal to 2.41 for 1970 and 2.36 for $2003 \div 2006$.

Conclusions

1. Power function dependence in Zipf law realization allows to conclude that popular regularities in economics (zipfian and also logarithmic) can have the common stochastic origin.
2. Zipfian behaviour is encountered also in chaotic dynamical systems with multiple agents (attractors). Deviations from linear dependencies in log-log scale allows to state the presence of perturbations.
3. Time-dependent or parameter-dependent exponent dynamics (from Zipf-Mandelbrot and Yule dependencies) allow to model and to estimate the survivor of economic system (self-organised criticality).
4. Cities are on average more evenly distributed than suggested by Zipf law.

References

1. G. K. Zipf. Human Behavior and the Principle of Least Effort. – Addison-Wesley, 1965.
2. Mandelbrot Benoit. Information Theory and Psycholinguistics. – In: R. C. Oldfield and J.C. Marchall. Language. – Penguin Books, 1968.
3. Simon H. A. On a class of skew distribution functions. – *Biometrika* 42(3–4) (1955) 425–440.
4. Gabriel Altmann. Prolegomena to Menzerath's law. – *Glottometrika* 2 (1980) 1–10.

5. Gibrat R. Les Inégalités économiques. – Paris, 1931.
6. Schorlemmer D., S. Wiemer, M. Wyss. Variations in earthquake-size distribution across different stress regimes. – *Nature* 437 (2005) 539–542.
7. Inmaculada B. Aban, Mark M. Meerschaert. Shifted Hill's estimator for heavy tails. – *Commun. Statist. — Simula* 30(4) (2001) 949–962.
8. M. Ausloos, N. Vandewalle, Ph. Boveroux, A. Minguet, K. Ivanova. Applications of statistical physics to economic and financial topics. – *Physica A* 274 (1999) 229-240.
9. Ian Wright. The social architecture of capitalism. – *Physica A* 346 (2005) 589–620.
10. D. Costantini, S. Donadio, U. Garibaldi, P. Viarengo. Herding and clustering: Ewens vs. Simon–Yule models. – *Physica A* 355 (2005) 224–231.
11. Rozenfeld H., Rybski D., Andrade J. S., Batty M., Stanley H. E., Makse H. A. Laws of Population Growth. – *Proc. Nat. Acad. Sci.* 105 (2008) 18702–18707.
12. Edoardo Galeo, Mauro Gallegati, Antonio Palestrini. On the size distribution of firms: additional evidence from the G7 countries. – *Physica A* 324 (2003) 117 – 123.
13. Robert L. Axtell. Zipf Distribution of U.S. Firm Sizes. – *Science* 293 (2001) 1818-1820.
14. Matthieu Wyart, Jean-Philippe Bouchaud. Statistical models for company growth. – *Physica A* 326 (2003) 241-255.
15. Luıs A. Nunes Amaral, Sergey V. Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael A. Salinger, H. Eugene Stanley, Michael H.R. Stanley. Scaling Behavior in Economics: I. Empirical Results for Company Growth. – *J. Phys. I France* 7 (1997) 621–633.
16. Yoshi Fujiwara, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, Wataru Souma. Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. – *Physica A* 335 (2004) 197-216.
17. Yoshi Fujiwara. Zipf law in firms bankruptcy. – *Physica A* 337 (2004) 219-230.
18. Yoshi Fujiwara, Hideaki Aoyama, Corrado Di Guilmi, WataruSou ma, Mauro Gallegati. Gibrat and Pareto–Zipf revisited with European firms. – *Physica A* 344 (2004) 112–116.
19. Jianhua Zhang, Qinghua Chen, Yougui Wang. Zipf distribution in top Chinese firms and an economic explanation. – *Physica A* 388 (2009) 2020-2024.
20. Hari M. Gupta, Jose R. Campanha, Daniela R. de Aguiar, Gabriel A. Queiroz, Charu G. Raheja. Gradually truncated log-normal in USA publicly traded firm size distribution. – *Physica A* 375 (2007) 643-650.
21. Byoung Hee Hong, Kyoung Eun Lee, Jae Woo Lee. Power law in firms bankruptcy. – *Physics Letters A* 361 (2007) 6–8.
22. Wataru Souma. Physics of Personal Income. 2002. – <http://arxiv.org/pdf/cond-mat/0202388.pdf>, accessed 2012.01.06.
23. Pareto Vilfredo. Cours d'Économie Politique: Nouvelle édition par G.-H. Bousquet et G. Busino. – Geneva: Librairie Droz, 1964. – Pp. 299-345.
24. British income taxes // In: Charles Gide. Cours d'économie politique. Vol. 1. – Paris, 1919.
25. Geza Hegyi, Zoltan Neda, Maria Augusta Santos. Wealth distribution and Pareto's law in the Hungarian medieval society. – *Physica A* 380 (2007) 271–277.
26. J.R. Iglesias, S. Gon,calves, G. Abramson, J.L. Vega. Correlation between risk aversion and wealth distribution. – *Physica A* 342 (2004) 186-192.
27. Adrian Dragulescu, Victor M. Yakovenko. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. – *Physica A* 299 (2001) 213-221.
28. R. Hernandez-Perez, F. Angulo-Brown, Dionisio Tun. Company size distribution for developing countries. – *Physica A* 359 (2006) 607-618.
29. J.J. Ramsden , Gy. Kiss-Haypal. Company size distribution in different countries. – *Physica A: Statistical Mechanics and its Applications* 277(1–2) (2000) 220-227.
30. Jesse H. Ausubel. Will the rest of the world live like America? – *Technology in Society* 26 (2004) 343–360.
31. Richard Trigaux. The wealth repartition law in an altruistic society. – *Physica A* 348 (2005) 453–464.
32. K. Okuyamaa , M. Takayasub , H. Takayasuc; Zipf's law in income distribution of companies. – *Physica A* 269 (1999) 125-131.
33. Hideaki Aoyama, Yoshi Fujiwara, Wataru Souma. Kinematics and dynamics of Pareto–Zipf's law and Gibrat's law. – *Physica A* 344 (2004) 117–121.
34. Atushi Ishikawa. Pareto law and Pareto index in the income distribution of Japanese companies. – *Physica A* 349 (2005) 597–608.
35. Amit K. Chattopadhyay, Sushanta K. Mallick. Income distribution dependence of poverty measure: A theoretical analysis. – *Physica A* 377 (2007) 241–252.
36. Nanak Kakwani. On a class of poverty measures. – *Econometrica* 48(2) (1980) 437-446.
37. Amartya Sen. Poverty: an ordinal approach to measurement. – *Econometrica* 44 (1976) 219-231.
38. Mao-Bin Hu, Rui Jiang, Qing-Song Wu, Yong-Hong Wu. Simulating the wealth distribution with a Richest-Following strategy on scale-free network. – *Physica A* 381 (2007) 467–472.
39. M. Naldi. Concentration indices and Zipf's law. – *Economics Letters* 78 (2003) 329–334.
40. Catherine Liston-Hayes, Alan Pilkington. Inventive Concentration: An Analysis of Fuel Cell patents. – *Science and Public Policy* 31(1) (2004) 15-25.
41. M. Ausloos, Ph. Bronlet. Strategy for investments from Zipf law(s). – *Physica A* 324 (2003) 30-37.

42. M. Ausloos, K. Ivanova. Precise (m; k)-Zipf diagram analysis of mathematical and financial time series when m = 6, k = 2. – *Physica A* 270 (1999) 526-542.
43. Carlos Alegria, Klaus Schaeck. On measuring concentration in banking systems. – *Finance Research Letters* 5 (2008) 59-67.
44. Thomas Wilhelm, Peter Hanggi. Power-law distributions resulting from finite resources. – *Physica A* 329 (2003) 499-508.
45. M. Ali Saif, Prashant M. Gade. Emergence of power-law in a market with mixed models. – *Physica A* 384 (2007) 448–456.
46. P.V. (Sundar) Balakrishnan, James M. Miller, S. Gowri Shankar. Power law and evolutionary trends in stock markets. – *Economics Letters* 98 (2008) 194–200.
47. Caglar Tuncay, Dietrich Stauffer. Power laws and Gaussians for stock market fluctuations. – *Physica A* 374 (2007) 325–330.
48. Roberto Reno, Rosario Rizza. Is volatility lognormal? Evidence from Italian futures. – *Physica A* 322 (2003) 620-628.
49. Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, H. Eugene Stanley. Institutional investors and stock market volatility. – *The Quarterly Journal of Economics* 5(2006) 461-504.
50. Arnab Chatterjee, Bikas K. Chakrabarti. Ideal-gas-like market models with savings: Quenched and annealed cases. – *Physica A* 382 (2007) 36–41.
51. Arnab Chatterjee, Bikas K. Chakrabarti, S.S. Manna. Pareto law in a kinetic model of market with random saving propensity. – *Physica A* 335 (2004) 155-163.
52. Pratip Bhattacharyya, Arnab Chatterjee, Bikas K. Chakrabarti. A common mode of origin of power laws in models of market and earthquake. – *Physica A* 381 (2007) 377-382.
53. Yoram Louzoun, Sorin Solomon. Volatility driven market in a generalized Lotka–Voltera formalism. – *Physica A* 302 (2001) 220-233.
54. Sorin Solomon, Peter Richmond. Power laws of wealth, market order volumes and market returns. – *Physica A* 299 (2001) 188–197.
55. A.M. Chebotarev. On stable Pareto laws in a hierarchical model of economy. – *Physica A* 373 (2007) 541–559.
56. P.K. Mohanty. Why only few are so successful? – *Physica A* 384 (2007) 75–79.
57. Zoltan Kuscsik, Denis Horvath, Martin Gmitra. The critical properties of the agent-based model with environmental–economic interactions. – *Physica A* 379 (2007) 199–206.
58. Tatsuo Yanagita, Tamotsu Onozaki. Dynamics of market structure driven by the degree of consumer's rationality. – *Physica A* 389 (2010) 1041-1054.
59. Volker Nitsch. Zipf zipped. – *Journal of Urban Economics* 57 (2005) 86–100.
60. Matteo Marsili, Yi-Cheng Zhang. Interacting Individuals Leading to Zipf's Law. – *Physical Review Letters* 80(12) (1998) 2741-2744.
61. Lucien Benguigui, Efrat Blumenfeld-Lieberthal. The temporal evolution of the city size distribution. – *Physica A* 388 (2009) 1187-1195.
62. Kwok Tong Soo. Zipf's Law for cities: a cross-country investigation. – *Regional Science and Urban Economics* 35 (2005) 239–263.
63. Juan-Carlos Cordoba. On the distribution of city sizes. – *Journal of Urban Economics* 63 (2008) 177–197.
64. Xavier Gabaix. Power Laws in Economics and Finance. – *Annual Review of Economics* 1 (2009) 255–93
65. Lucien Benguigui, Efrat Blumenfeld-Lieberthal. A dynamic model for city size distribution beyond Zipf's law. – *Physica A* 384 (2007) 613–627.
66. Rui Carvalho, Alan Penn. Scalingand universality in the micro-structure of urban space. – *Physica A* 332 (2004) 539 – 547.
67. D. Volchenkov, Ph. Blanchard. Scaling and universality in city space syntax: Between Zipf and Matthew. – *Physica A* 387 (2008) 2353–2364.
68. Damian H. Zanette, Susanna C. Manrubia. Role of Intermittency in Urban Development: A Model of Large-Scale City Formation. – *Physical Review Letters* 79(3) (1997) 523-526.
69. Gilles Duranton. Some foundations for Zipf 's law: Product proliferation and local spillovers. – *Regional Science and Urban Economics* 36 (2006) 542–563.
70. Yuri Mansury, Laszlo Gulyas. The emergence of Zipf's Law in a system of cities: An agent-based simulation approach. – *Journal of Economic Dynamics & Control* 31 (2007) 2438–2460.
71. Ethan H. Decker, Andrew J. Kerkhoff, Melanie E. Moses. Global Patterns of City Size Distributions and Their Fundamental Drivers. – *PLoS ONE* 9 (2007) e934. – <www.plosone.org>, accessed 2010.06.04.
72. Ferdinando Semboloni, Francois Leyvraz. Size and resources driven migration resulting in a power-law distribution of cities. – *Physica A* 352 (2005) 612–628.
73. Jose Maria Sarabia, Faustino Prieto. The Pareto-positive stable distribution: A new descriptive model for city size data. – *Physica A* 388 (2009) 4179-4191.
74. Gordon Anderson, T. Ying Ge. The size distribution of Chinese cities. – *Regional Science and Urban Economics* 35 (2005) 756–776.
75. Kausik Gangopadhyay, B. Basu. City size distributions for India and China. – *Physica A* 388 (2009) 2682-2688.
76. Hsin-Ping Chen. Path-dependent processes and the emergence of the rank size rule. – *Ann. Reg. Sci.* 38 (2004) 433–449.
77. Newton J. Moura Jr., Marcelo B. Ribeiro. Zipf law for Brazilian cities. – *Physica A* 367 (2006) 441–448.