TRENDS IN DATA ANALISYS: STATE, DEVELOPMENT PROSPECTS

Doctor in Economic sciences, Prof.,I. A. Katsko; PhD in Economic sciences, P. Yu. Velichko;
Student,M. Nikogda
Kuban State Agrarian University – Russia, e-mail: stat@kubsau.ru

On the banks of the Rhine for many centuries was towering beautiful castle. The spiders, which dwelt in the cellars of the castle, tightened all its aisles with cobwebs. Once a strong gust of wind destroyed the thinnest threads of the web, and the spiders began to recover the gaps: they believed that the lock was kept on their web!

M. Klain [10]

Abstract*. The article suggests the consideration of data analysis ideology in the context of knowledge creation process and the technological patterns of social development. The problems of singularity (human misunderstanding of data processing results) associated with increase in data variety, volume and further intellectualization of the corresponding technologies for their processing are proposed to be solved by creating new formalization techniques that allow retransmission.*

KEY WORDS: *DATA ANALYSIS, TECHNOLOGICAL PARADIGM, ANALYTICS, KNOWLEDGE, MACHINE LEARNING, BIG DATA AND THE INTERNET OF THINGS, MEGATRENDS, AMICABLE INTELLIGENCE OF THE HUMAN LEVEL, FORMALIZATION, CONSTRUCT, SCRIP*

1.Introduction

Data analysis (applied statistics) as development of statistics ideology, probability theory and mathematical statistics intensively developed over the last two centuries is naturally considered in the context of the socio-economic development of society with regard to solution of management and decision-making problems [16-18]. In this case, it seems interesting as a context to lead the ideology of technological paradigm[1], suggested by D.S. Lvov, S.Yu. Glazyev, G.G. Fetisov and essentially relied on larger cycles by N.D. Kondratiev (the phase of new ideas emergence - lasts about 10 years, the phase of paradigm growth - does about 40 years, the maturity phase lasts about 10 years more) [4, 5, 12]. Using the data analysis as an example, it is easy enough to trace the tendencies of the decision-making support ideology that are based on empirical observations characteristic for one or another way (Table 1)

2. Data analysis – contextual approach.

In postindustrial society, the cognitive revolution, which began in the 1950s and 1960s, manifested itself. They can talk about two periods of its development. I cognitive revolution, where a person is a carrier and a generator of knowledge, and a computer, the Internet and software are tools based on the machine learning ideology (the artificial intelligence implementation in a weak version - machine intelligence). II cognitive revolution - new knowledge is generated by the computer (the artificial intelligence formation).

The first scientific paradigms had a material basis, which is very important for the social adaptation of man in the real world. Cognitive paradigm, based on machine intelligence (and in the long term amicable *artificial intelligence of the human level - AIHL* (DIYCH-in Russian), called the fourth industrial revolution, is focused on accelerating all processes by integration at the expense of information technology and the Internet of all things, the basic megatrends of modern society (physical, digital and biological) [1, 19, 22-24].

The purpose of data analysis is to obtain new knowledge about the studied system using observations or differently to convolve (compress) existing information for solving applied problems of analysis and explaining the features of the studied system functioning , management, forecasting (predication) and decision-making.

The main difference between applied statistics (data analysis) from mathematical statistics is the consideration of not only probabilistic but also geometric and logical nature of data, as well as the obtaining of convolutions by both formal algorithmic methods (classical methods of *multidimensional statistical analysis* - MSA) and not formal ones (machine learning and adapted methods of MSA).

According to E. Toffler's studies, nowadays, power in society is based on three basic elements: strength, money and knowledge. [21] Moreover, knowledge becomes a universal tool that can replace all others. This is why Russell Ackoff's definition, which characterizes the process of the knowledge formation, acquires a special meaning, which in our formulation is expressed in the following way [2]

Facts – Iformation – Data – Knowledge – Understanding – Wisdom.

One of the forms of knowledge representation contributing to thinking formation and worldview of human has always been mathematics, which allowed to form a chain of thinking levels (*recognition - reproduction of model situations - atypical situations analysis - creativity*). The most important stage in the process of knowledge formation is understanding - a person easily perceives and uses in practice what is understandable. From the point of view of data analysis, the stages of the knowledge formation can be disclosed in the following way [2, 6]:

- *facts – events, that have already happened;*
- *information – facts characteristic;*
- *data – facts, described quantitatively or qualitatively, presented in the form of tables «object – property (feature)» or «question - answer»;*
- *knowledge – rules «If …, so …», which can be used in decision-making;*
- *understanding – presentation about functional features of studied object, managing possibilities,foresight (predication) and decision-making;*
- *wisdom – ability to use the reached understanding in future.*

Table 1 *– Data analysis in context of socio-economical development*

Techno logical way	*Main formalization form (approach*	*Data processing way*
I mechanizat ion 1770-1830 II steam engines, railways 1830-1880 III electricity, metallurgy 1880-1930	Mathematical analysis (data – realization results of mathematical laws of natural sciences)	Descriptive statistics, differential and integral calculus
IV oil, mass production, nuclear power 1930-1980	Probability theory and mathematical statistics (data – random processes realization, which submit to certain distribution laws – parametrical statistics, or nonparametrical statistics)	Selective method, Convolution of information. Formalistic algorithmic approaches, solving problems: Data description, visualisation, classifications and dimension decrease, search of dependences (multidimensional statistical analysis)
V informatiza tion, telecommu nication 1980-2020	Data analysis (applied statistics) (any data nature; probabilistic, geometric, - data form in multidimensional attribute space «compact» (clots), logical – this not only quantitative, but also non-numerical (qualitative) form patterns – interrelations not always explainable at the quantitative level)	Analytics 1.0 (descriptive analytics) OLAP cubes, convolution procedures that do not allow an algorithmic approach - Exploratory data analysis (EDA) , also based on the computer training ideology (Data Mining) as an option for implementing EDA based on information technology, web-sites scraping
VI nano-, bio-, info-, cognitive-, socio-technology – *NBICS* 2020-2060	*Big Data, Internet of Things* (*IoT*) (data nature is any of the above, including visual, textual, sound, video-audiofiles and others)	Analytics 2.0 (predictive analytics) Analytics 3.0 (prescriptive analytics - the basis of *CRM*) Analitics *N.0* (analitics, supporting typical solutions, having opportunities to search and process information on-line/interface (analogue of modern app *Siri* from *Apple*)
VII human – main technology subject 2060-2100	... Data nature is any of the above, including not representable in semiotics systems projection foresight	Analytics *NBICS, based on technologies of amicable intelligence of human level*

In order to use methods of applied statistics, the data must be measured with qualitative or quantitative scales. The measurement process is accompanied by problems: heterogeneity, quality, limitations, subjectivity of perception and thinking. Moreover, person is limited in perception of the surrounding world. As it is known, it is characterized by the number of J. Miller (1956) (7 ± 2) - according to it, there is a need to compress large volumes of information and representation in the form of (preferably understandable) models. This goal is devoted to the work of decision support systems that allow solving tasks: descriptive statistics (*OLAP* cubes), classification and diminution of dimensions, search for dependencies, prediction, etc. implemented in *KDD* class systems and *Data Mining.*

The implementation of machine learning methods does not allow us to realize the "understanding" stage in R.Akoff's knowledge formation process and shows the practical application of Braimean's uncertainty principle, which is an analogue of Heisenberg's uncertainty principle in data analysis context:

«*accuracy* × *interpretability* =
= *Braimean's constant*».

The famous futurist E. Toffler talks about three waves in the development of society: agrarian, industrial, informational, which their traditional education systems conformed to [19]. For several decades, we have witnessed the transformation of education traditional system for an industrial society. Data analysis has always evolved in the direction of meeting the needs of society. If there was enough descriptive statistics in the agrarian society, in industrial - analytical statistics, the postindustrial society and the expected information extended the applied statistics with machine learning methods using both structured and unstructured data (Data Mining, Text Mining, Web Mining, Social Mining , Big Data and the Internet of things), thus, the demand for data analysis methods is determined by the level of development of the society, for example, business intelligence technologies that have long been used in Moscow and St. Petersburg, are developing in the regions only in recent years, becoming one of the costly business articles (Table 1). The development of the digital sector of the economy is being discussed in the world, therefore, it becomes necessary to systematically comprehend the possibilities, limitations and prospects for data analysis at the present stage of society development.

3. ***Megatrends.***

Today's data processing methods are based on the ideology of probability, statistics, data analysis and, among other tasks, allow solving the problem of finding "megatrends" (conditional coordinate system) at different levels of society at a new level that allows explaining many phenomena in the socio-economic space. The number of "megatrends" for a society is the same as for a person (7 ± 2) and corresponds to one of the theories claiming the theory of "great unification" in physics - superstring theory, which requires about 10 measurements. At present, one of the obvious "megatrends" explaining the transformation of the education system is that "the transition to an information society requires getting rid of the outdated educational system that trained cadres for the industrial society." Analysis of printed sources allows us today to talk about the following "megatrends" of our society, conditioned by the digital revolution and carrying a disruptive influence (a form of natural selection in biology), "tearing" the homogeneous aggregate (usually) into two extreme variants and not contributing to the average state, which explains the appearance in the socio-economic systems of power law distribution laws, such as the Pareto or Zipf law (for example, the result of such an impact can be considered the stratification of society: rich and poor, "golden" Billion and others, etc.) [1, 2, 8, 9, 19-24] :

1) Mankind today lives in a consumer society and gradually loses the distinction between the real and the virtual.
2) Priority in life is obtained by "physical" people, without moral and moral obligations.
3) The society increasingly depends on information technology (3D-printing, development of 4D-printing

technology, the production of any services and goods in online mode).

4) Information technology is becoming one of the subjects of everyday life of modern man (unmanned vehicles, robotics, new materials).
5) People pay and receive money for virtual actions that do not give a sense of physical incarnation, which negatively affects the human psyche.
6) New approaches to interaction and cooperation at all levels of the society are being developed. For example, "distributed databases" are *block chains* that represent a data store that is available for verification to anyone (for example, *Bitcoin*).
7) Particular attention is paid to *Big Data* concepts and the "*Internet of all things*" in the study of social networks, industry, business.
8) The growing opportunities for biological engineering require the development of a normative, ethical and legal framework.
9) The transition to an information society requires transforming an outdated education system that trained personnel for an industrial society by selecting a goal at the state level (for example, harmonious development of a person) instead of replenishing a certain labor market (or "human cloud").
10) The actual sector of the economy, based on the "human cloud", is becoming relevant.

Modern information technologies for data analysis (*web mining and text mining*, etc.) make it possible to find an alternative to classical content analysis when searching for "megatrends" including without human participation (scraping *web*-sites).

"Megatrends" change over time and the past ("megatrends") "twists" along the new ones, relying on the general direction (mainstream) of the 21st century - a digital revolution that, like the communist movement 100 years ago, will change the landscape of the planet. Perhaps right now there is a tectonic gap between the "golden billion" and the rest of the world's population, although the maturing changes are not universal even for developed countries. For example, the politics of the consumer society is unacceptable in the Arab world and can have a different form (as in India and China). In Russia there is a centro-peripheral model of socio-economic space (N. Zubarevich) [8]:

- post industrial Russia (federal cities with a million population with postindustrial economy),
- industrial Russia (industrial cities with a population of up to 250 thousand people),
- rural periphery
- agrarian Russia (the main part of the country and residents of settlements with a population of less than 20 thousand people) ,
- patriarchal republics, based on their own values (the North Caucasus, Southern Siberia).

And all four of Russia perceive different future changes and react differently.

At different levels of the hierarchy of society, there are their own tendencies to change the world, so at the state level (in most countries) first of all one can distinguish: bitcoin, crypto-currencies, cyberwar, fakes.

4. The problem of translation of knowledge.

Current trends in the development of analytics are directly related to the achievements of human intellect, which cause a futurist (fear of the future). Mankind is preoccupied with its own ideas about the power of computers and the alleged consequences of the emergence of artificial intelligence (AI). Many scientists expect the emergence of the point of "singularity" - the moment when the possibilities and results of the activity of artificial intelligence systems in the narrow sense (understood as the realization of the ideology of machine learning) will surpass the possibilities of human understanding [1, 9]. In addition, artificial intelligence is expected to reach the human level. The only thing people hope for is that they will be a *friendly human intellect (AIHL).*

Intelligence of information systems is achieved today through the use of the Internet (the Internet of things), the ideology of machine learning, and computational capabilities. Thus, we are not talking about the presence of consciousness, but it is possible that the opportunity will come of "creating the effect of consciousness" through the effect of replacing the computational abilities. Then only the initial education of unconditional friendliness to man will pass the point of singularity.

In fact, we are talking about machine intelligence, which, due to the processing power, can generate new knowledge in the form of patterns and (or) constructs that can not be explained by humans on the basis of available information (including using *Big Data* and the Internet of all things). The traditional sequence of the process of forming knowledge according to R. Akoff (Facts - Information - Data - Knowledge - Understanding - Wisdom) will be broken. For if traditionally for a person of "knowledge" are products (rules) "if ..., then ..." that allow to realize the stage of "understanding", then the question arises about the need for a new round of development of mathematics and information technologies oriented to "not ... exclusion "of man from the processes of management and decision-making in the socio-economic space, since the knowledge obtained by the *AIHL* can go beyond the boundaries of human understanding (the rules" if ..., then ... "). Data analysis today is a way of compressing large volumes of information to support decision-making processes, which allows to identify patterns in data and to present them in the form of: graphs, tables, formulas, various dependencies obtained using machine learning methods. It is assumed that at the point of singularity, knowledge will go beyond these limits.

The problem of understanding the intellectual systems of the future is similar to the problems of the middle of the last century, when the possibilities of interaction with computers were formed through programming languages, which today number more than 8000.

We believe that in order to solve the problem of "retransmission" of knowledge to a person, the potential capabilities of the AIHL must presuppose the possibility of synthesizing a number of subject areas (SA) into which new knowledge can be projected. The lexicographic ordering of the SA will allow to identify and rank the consequences of applying new knowledge in different areas.

Thus, it becomes urgent to develop an understandable ideology of the description of the SA. As a possible example, consider, following the work of L.S. Bolotova [3], the method of situational analysis and design of the design of the SA model, on the basis of the set-theoretical (relational) approach, in the form of a complex of invariant constructs as applied to the description of the SA for the new knowledge.

4.1. The domain model. The synthesized object (system) should be created under the condition of the existence of an external environment that is characterized by a certain subject area (SA) - a part of the real world within the given context (industrial, agricultural, financial, computer, etc. corresponding to the direction of knowledge). Each subject area has its own language, which can be formalized using binary relations [3]. Usually, a system is understood as the set of a related set of objects. Most often, two sides of connectivity are considered: as a fact of the existence of a relationship between individual elements of the system - realizing the cognitive conceptual aspect (cognitive maps); as a description of the process of the corresponding connectivity of elements - a functional, information or behavioural aspect (semantic networks, frames, products, methods of situational modelling). Both approaches are considered rather rarely.

Modelling *SA* of an arbitrary nature is connected, first of all, with the analysis of the categories describing it. A category is understood as a construct or otherwise some abstract container, with some objects entering into it, and others not (this is the postulate of

our thinking for more than two millennia). It is assumed that the categories that a person operates on can be arranged in the following hierarchy: the higher level - the base level - the lower level. The base level is the level at which most of our knowledge is structured. It gives an opportunity to perceive geometric visualization of the conceptual structure of an object.

4.2. Algebraic models of SA. At present, algebraic language and style of thinking are the standard approach to the representation of data and knowledge in information systems. The question of the possibility of a correct description of the model of the subject domain associated with the person (observer) in the form of a system of objects with certain relations can be investigated only by means external to this system (that is, in some other theory), as follows from K. Gedel's theorem on incompleteness of formal arithmetic (to which almost all mathematical theories can be reduced).

D. Hilbert describes the interpretation of the formalization of mathematical theory, as well as the method that makes the formal system the subject of the study of mathematical discipline - metamathematics or the theory of evidence [11].

The introduction of a system of *S* objects according to the ideology of metamathematics, assuming the existence of a non-empty set of objects between which certain relations are established, can proceed from two methods characterizing two main trends in modern mathematics - constructivism (modern predecessor of which is A. Poincaré and which was basic in ancient science, for example , in Euclid) and formalism, which implies a complete abstraction from the meaning.

In an axiomatic method, the axioms underlying the formal approach are used as assumptions about the system of *S* objects. Then we examine the consequences of the axioms, which form a theory with respect to the system of S objects under consideration.

A constructive (genetic) method involves constructing objects in a certain order. S. Klini characterized this approach as a method of *substantive or material axiomatics*. To describe the system from the point of view of the observer observing the system from the outside, a "formal system object" is introduced-the meta-set of the research system S, which allows describing systems based on logical and mathematical methods [13].

The presence of experts with exact, technological, effective thinking, free "from traditions and cognitive prejudices", which will interact with *AIHL*, is postulated. To do this, special (psychological) work with experts and subjects of the problem of creating a new object is supposed, accustoming them to operate with their "constructs".

The basic representation of the construct is the meta-set

$$S<X_{as}, X_a, X_{ao}, \cup X_{ac_i}>,$$

where X_{as} – (*action subject*), X_a – (*action*), X_{ao} – (*action object*), $\cup X_{ac_i}$ – (*action components*). All elements of meta-set have (*property*):

$X_{as}= X_{as}$ (p_{s1}, …, p_{sl}), $X_a= X_a(p_{a1}, …, p_{aq})$, $X_{ao} = X_{ao}$ (p_{o1}, …, p_{ot}), $X_{ac_i} = X_{ac_i}(p_{ac_i1}, \dots, p_{ac_ih})$, the results of relations of which among themselves within the framework of our task realize TER (technological, technical, operational, economic, environmental requirements for the subject area, which are based on the normative provisions defined by the person).

A logical representation of a construct implies two components:

1) functional – Φ , regarded to the purposes of building a new object and described as a union of binary relations (*R*)

$$\Phi = R_{as}(X_{as}, X_a) \cup R_{ao}(X_a, X_{ao});$$

2) providing, (achievement of the goal) – Q,

$$Q = R_{as}(X_{as}, X_a) \cup R_{ac_i}(X_a, X_{ac_i}).$$

$$K = \Phi \cup Q .$$

Each construct *K* is a kind of domain concept that is open to expansion and modification, which is intended for reusable use in designing, obtaining production rules, and so on.

Combining all constructs:

$$U = \cup \mathrm{K_j}$$

gives us the universum *U*, a structure called a polyhedron in topology, describing an SA of arbitrary nature.

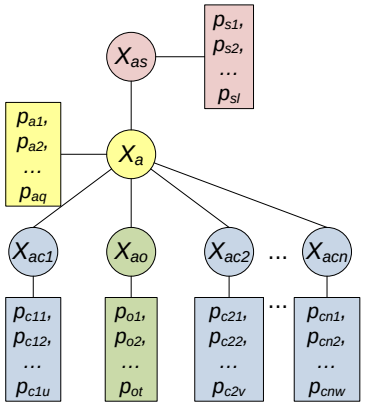

Figure 1 *– Construct model – Subject area extract [3]*

The universum is a generalized model of a specific subject area, which can be represented as a basis for concepts (*ontologies*) designed for reusable, multipurpose use in various applications and relationships between them that implement production rules [14]. Consideration of ontologies with selection functions and mechanisms for their implementation allows us to talk about a knowledge base that potentially allows the formation of products (rules) that are understandable to man [15].

5. Conclusions. The problems of human interaction and the human-level friendly intellect require the creation of means for "retransmitting" to a person new knowledge obtained by advanced methods of data analysis (structured, weakly structured, unstructured). At the author's level of vision, developments are required that allow translating, obtained knowledge (patterns, constructs, etc.) into an "understandable" kind of person, for example, projecting into several mutually complementary subject areas, which can be realized (*AIHL*) described by the method of situational analysis and design of the design of the SA model. Within the framework of the domain model - the creation of a new object (or description of the situation) is reduced to operations over the relations between constructs and their elements. Thus, one of the variants of "passing the point of singularity" is the formation of knowledge bases related to a certain subject area and *an assessment of the consequences of the implementation of new knowledge through the use of scenarios based on the SA model*.

The emergence of a new scientific paradigm in science and society forms a new space in which the previous includes (due to) a new generalization, the postulates (principles) change and shrink, with an increase in the coverage of the phenomena described (the scaling effect of socio-economic space) [7], which is realized in an explicit form using the example of data analysis. The further development of intelligent information systems leads to the automation of the work of analysts and other professionals, the emergence of new knowledge and the ability to present them for adequate human perception, and the need to create ethics councils, to address employment and social issues is already generally recognized. To preserve one's identity in the future information society, a person needs to solve many problems: preserve human culture, universal values, find for himself and implement new means of formalization (adapt or develop new mathematics for translation of knowledge obtained by the *AIHL*), etc. Everyone should understand that it is he who builds the future world and how he will solve it.

6. Literature

1. J. Barrat , The last invention of mankind: Artificial intellect and the end of the era of *Homo sapiens*. /from English. Natalia Lisova. - M.: Alpina non-fiction, 2015. - 304 p.
2. P.S. Bondarenko Theory of Probability and Mathematical Statistics: Textbook / P.S. Bondarenko, G.V. Gorelova, *I.A. Katsko*; Ed. by I.A. Katsko, A.I. Trubilina. Moscow: KNORUS, 2017. - 390 p.

3. *L.S. Bolotova*. Conceptual design of the domain model with the help of software systems for the development of knowledge bases for intelligent decision support systems / L.S. Bolotova, V.A. Smolyaninova, S.S. Smirnov // High technology: scientific. -technical. - 2009. - T. 10 No. 8. - P. 28-36.
4. S. Yu. Glazyev , D.S. Lvov, G.G. Fetisov. Evolution of technical and economic systems: the possibilities and boundaries of centralized regulation. - M.: Science. - 1992. - 208 p.
5. S.Yu. Glazyev. Strategy of advanced development of Russia in the conditions of global crisis. – M.: Economics, 2010. - 255 p.
6. N.G. Zagoruyko. Applied methods of data analysis and knowledge. – Novosibirsk, Uni of maths, 1999. – 270 p.
7. V. A. Zorich Mathematical analysis of natural science tasks – M.: MCNMO, 2017. – 160 p.
8. N. Zubarevich. «Four Russia-s" and new political reality. *Web: http://polit.ru/article/2016/01/17/four_russians*/
9. M. Kaku, Mind future / from eng. by N. Lisova, – M.: Alpina non-fiction, 2015. – 502 p.
10. M. Kline. Mathematics. Loss of certainty. - The World, 1984. - 134 p.
11. S. Klini. Introduction to metamathematics: - M.: Librocom, 2009. - 528 p.
12. N.D. Kondratiev, Large cycles of conjuncture and theory of foresight / edited by AL Albakina. – M.: "Publishing house" Economics ", 2002. - 767 p.
13. V.V. Kulba etc. Information processes and information management / Human factor in management сб. Articles of the ISP RAS ed. O.N. Abramova, K.S. Ginsberg, D.A. Novikov and others - M .: KomKniga, 2006. - 496 p.
14. Y. Lyapar. Theory of System-Structural Design - the Basis of Intellectualization of Modelling and Decision Support Systems / Yu.I. Lypar, I.A. Katsko, G.F. Bershitskaya - Krasnodar .: KSAU, 2010. – 49 p.
15. Yu. I. Lyapar. Synthesis of knowledge bases of analog electronic devices. // Ulyanovsk: Works of the Intern. Conf. "Continual and Algebraic Logic, Calculus and Neuroinformatics in Science and Technology", vol. 3, 2006, p.120-128.
16. A.V.Maltseva, N.E. Shilkina, O.V. Mahnitkina Data mining sociology: Experience and outlook for research // Social surveys 2016-January(3), p. 35-44.
17. D.A. Novikov. Cybernetics: Navigator. History of cybernetics, current state, development prospects. – M.: LENAND, 2016. - 160 p. (Smart Management Series)
18. A.I. Orlov. The main features of the new paradigm of mathematical statistics // Polytematic network electronic scientific journal of the Kuban State Agrarian University (Kuban State University, Krasnodar, KubSAU, 2013. № 06 (090)., P. 188-214 http://ej.kubagro.ru/
19. D. Rose, The Future of Things: How a Fairy Tale and Fantasy Become a Reality / 3rd edition / translation from the English Seeds of Sheshenin .- M.: Alpina non-fiction, 2017. - 344 pp.
20. E. Toffler. The Third Wave – M.: AST, 2010. - 784 pp.
21. E. Toffler. Metamorphoses of Power – M.: AST, 2004. - 672 pp.
22. M. Ford. Technologies, who will change the world. Alexandra Kardash, M.: Mann, Ivanov and Ferber, 2014. - 268 p.
23. M. Ford. Robots come: Technology development and the future without work / Translated from English by Sergey Chernin .- M .: Alpina non 2016. - 430 p.
24. K. Schwab. The Fourth Industrial Revolution / K. Schwab – M.: "E", 2017. - 208 p.