37TH INTERNATIONAL SCIENTIFIC CONFERENCE ON
ECONOMIC AND SOCIAL DEVELOPMENT –
"SOCIO ECONOMIC PROBLEMS OF SUSTAINABLE DEVELOPMENT"
BAKU, 14-15 FEBRUARY 2019

Faculté des sciences juridiques économiques et sociales-salé

CERTIFICATE OF APPRECIATION

HEREBY WE PROUDLY CONFIRM THAT

Aida Guliyeva

HAS SERVED AS A DELEGATE TO THE ESD CONFERENCE AND SUCCESSFULLY PRESENTED THE PAPER:

TARANA ALIYEVA, ULVIYYA RZAYEVA ■ BAYESIAN APPROACH TO THE REDUCTION OF INFORMATION ASYMMETRY IN ENTERPRISE

PROF. SAKIT YAGUBOV
SCIENTIFIC COMMITTEE

BAKU, AZERBAIJAN
14-15 FEBRUARY, 2019

BAYESIAN APPROACH TO THE REDUCTION OF INFORMATION ASYMMETRY IN ENTERPRISE

Tarana Aliyeva
Azerbaijan State University of Economics, Azerbaijan
tarana.aliyeva@unec.edu.az

Aida Guliyeva
Azerbaijan State University of Economics, Azerbaijan
aida.quliyeva@sabah.edu.az

Ulviyya Rzayeva
Azerbaijan State University of Economics, Azerbaijan
ulviyya.rzayeva@unec.edu.az

ABSTRACT

Nowadays the processes of making and implementing managerial decisions are complicated by the asymmetry of information or its incompleteness. The article explores in detail the problem of information asymmetry in the enterprise and its consequences and gives recommendations on the practical application of information technologies for its elimination based on the Bayesian approach.

Keywords: *Decision making in uncertain states, Information asymmetry, Naive Bayes algorithm*

1. INTRODUCTION

Today the amount of information grows with a huge speed, but processes of acceptance and implementation of administrative decisions become complicated in view of its asymmetry. Big data technologies create a possibility of "special inquiries" for the purpose of elimination of information incompleteness and increase its transparency. It becomes possible when connections between productions, supply and demand are formed as result of identification of the hidden relations, templates and tendencies in data. The purpose of work is the research of an ability to apply new technologies to the management functionality associated with the processing of large volumes of information. Achievement of this goal requires the solution of the following tasks:

1. To investigate methods of receiving and aggregation of data from different sources, mutual and expeditious exchange of reliable information between organizational structures. All these processes require the development of analytical tools in the enterprise.
2. Taking into account the possibility of variants for situations with information asymmetry in making complex management decisions, to analyze the consequences of the problem of information uncertainty and incompleteness, considering these decisions as a concept of strategic orientation and an argument in favor of increasing the information transparency of the company.
3. To investigate the use of information technologies in the protection of confidential data, forcing enterprises to analyze more thoroughly the issue of ensuring the safety of information.
4. To investigate the specific consequences of information asymmetry in the enterprise and provide recommendations on the practical application of information technologies on its reduction by Bayesian method.

2. LITERATURE REVIEW

One of the negative consequences of high-intensity data circulation at enterprises Nikitina (Nikitina, 2011, pp. 61-65) calls information uncertainty situations arising through explosive growth of the information amount; while the task of monitoring and regulating the processes of achievement the set goals is being formed (Akerlof, 1970, pp. 488-460). In most cases decision-making is carried out in the conditions of a lack of information therefore, the qualitative analytics without qualitative knowledge is impossible. Paraphrasing the well-known phrase of Archimedes "Give me a point of support, and I will move heaven and earth", Maier-Shenberger (Mayer-Shenberger, Kuker, 2014) accents that today the huge amounts of data having an opportunity to transform all world data volume. The intensification and globalization of information streams has allowed identifying links and finding semantics in the integrated data, which do not have any sense, whether they are separated. Daas (Daas, Puts, 2014, pp. 22-31) calls such situation "mosaic" effect. Holmstrom at (Holmstrom, Milgrom, 1991, pp. 24-52) writes that the quality of the made decision quite strongly is determined by degree of readiness for information providing. Kudzh (Kudzh, 2016, pp. 23-27) connects a problem of such information asymmetry with complexity of processing of information's large volumes for expert estimation. Zhou (Zhou, 2013, pp. 24–25) sees the solution of this task in providing experts with the most objective data on a problem with the help of probabilistic methods. The relationship between data and their interdependence are inherent to the internal logic of large volumes of data. It allows finding nonconventional ways for the solution of planning problems and methods of plans' updating even after their acceptance, making macroforecasts without prejudice to primary activity in connection with temporary logs and strengthening control in real time (Liu, 2015, pp. 57–66). In this paper, an example of processing of hypotheses by means of Bayesian strategy will be reviewed. This strategy will be used to make a decision after processing information by certain algorithm. Considering a special role, which data play in education, our example concerns the information processing circulating in high school. These data accumulate a huge number of information on the person, since his deep childhood: this information includes data obtained in kindergartens, schools, colleges and universities, in institutions of professional development and vocational courses. The example considers only one aspect of the information asymmetry at university, concerning the employment of future graduates, since the processing of the person's formation involves a training data set, containing several hundred thousand elements and a large number of features (Vasyutinskaya, 2016, pp. 14-20). In the example, the Bayesian algorithm is simple and extremely convenient for working with large data sets, the advantage of which lies the assumption of features' independence (Foreman, 2016).

3. RESEARCH METHODOLOGY

In the solution of specific management problems, the creation and practical implementation of new methods and approaches to the development of socio-economic projects and programs becomes of particular urgency. To achieve this goal, the following tasks were set and solved in this paper:

1. The presence of huge amounts of unstructured information in the enterprise and the impossibility of its processing by standard means contributes to the concealment or loss of useful, valuable information when making managerial decisions. The paper explores how new technologies can meet complex management needs to eliminate problems caused by the circulation of information in the enterprise, thereby reducing its asymmetry.
2. Today, information in the business environment is considered not as a service, but as a product with all its inherent properties, advantages and disadvantages. The problem of information asymmetry of the enterprise and its consequences are investigated.

3. The organization's information security necessarily provides the protection of confidential data. It is shown that the provision of reliable information security in the information system involves the introduction of modern technology products to protect personal data and confidential information.
4. Informational asymmetry can be reflected in various facts and trends taking place in the enterprise. The specific consequences of information asymmetry are investigated and recommendations are given on the big data practical application to eliminate it.

4. RESULTS

Informatization in the enterprise management system generates a huge amount of extremely diverse and rapidly growing data; requests for their processing give impetus to the development of new analytical platforms. These changes are the main reason for the surge in the popularity of the big data concept. However, big data consist primarily of "raw" information, to find meaning in them; advanced technological resources of business management are needed. We will notice that asymmetry of information is shown both in its distribution, and in its acquisition – in our digital age there are unlimited opportunities for information search, but at the same time, there are no guarantees for its reliability. The fact that unstructured information increases the threat of data confidentiality in the enterprise, because data process arrays of complex, heterogeneous, uncertain nature, and spontaneous combinations of these data may contain certain characteristics that facilitate the disclosure of private information. This is an important issue when distributing microdata arrays from big data sources. Modern high-tech information products reliably protect personal data and confidential information for large-scale calculations in any distributed environment (in parallel computing systems, multiprocessor architectures, multithreaded, multitasking technologies, etc.) (Muller, Stallard, Warren, 1996). A distinctive feature of these technologies is a flexible response to changes in the entire IT infrastructure of the enterprise (adding new versions, functions and tools to an existing platform). In addition to high-tech products, national statistical offices and international organizations have clearly defined rules, which are accurately determined by the legislation when collecting, processing, analysis, distribution, preservation, protection and use of information in the activity. These laws and the corresponding strategies are designed to provide the rights of citizens for protection of personal information confidentiality, promoting thereby strengthening of their trust to the government and private institutions. In our article, we model a competitive job market, in particular, the change in market outcomes after the use of big data. We note that there are many uncertainties about the future and the impact of data on the job market (Widenhofer, Ytterstad, 2017). Therefore, we are building a flexible structure that covers a wide range of plausible results. In the study of complex social and economic systems, which require a high degree of uncertainty or information asymmetry Bayesian approach is very perspective. In the proposed paper, a mathematical apparatus is developed on the base of naive Bayes algorithm (NBA) that allows combining the available a priori ideas about the object with selective information with the assumption of signs' independence (Aliyeva, 2015, 257-262). In reality, sets of completely independent signs are extremely rare. Under an unlimited increase of sample size, Bayesian estimates coincide with classical estimates. That is why NBA gives a more effective result when applying to decision-making tasks. In our model, we assume that we have a set of university graduates with a density of 0 to 1, belonging to two categories - those who can expect to find work faster (experience, knowledge, assessments, etc.) and all others. We denote these categories by $k=H, L$, respectively. Employers often classify graduates into different segments based on various observable criteria. In these segments, we believe that there are still differences in intellectual, business, qualification and other status, which are explained by more intangible and unobservable factors. Therefore, our model will allow us to analyze these segments, and then we can think of our graduates as a specific segment of applicants.

We denote the number of graduates with a low employment rate through θ, while the remaining people with a high employment rate are 1-θ. Note that the parameter θ can vary. In addition, we assume that there are two types of graduates in each category: with a high probability of acceptance to one or another group, and with a low probability, which make up four groups. Next, we determine the probability of investment's loss in the formation of a particular student D as p^i, $i = l, h$, where $p^h > p^l$, which reflects individual risk. We denote the number of persons with low probability in a group with a low coefficient as t_L, while the number of persons with low probability in the group with a high coefficient is denoted as t_H. The total amount of individuals with a low level of probability is equal $\tau = t_L + t_H \leq 1$, and high probabilities will be $1 - \theta + \theta - t_L - t_H = 1 - \tau$. Then we can determine the average probability of employment across the population as $\bar{p} = \tau p^l + (1 - \tau)p^h$. The average probability of employment for the group with a low coefficient is $\bar{p}^L = (t_L/\theta)p^l + ((\theta - t_L)/\theta)p^h$, and the average probability of employment for the group with a high coefficient is $\bar{p}^H = (t_H/(1-\theta))p^l + ((1-\theta-t_H)/(1-\theta))p^h$. We determine the social gradient of the coefficient when the proportion of persons with a low probability level is higher for the group with a higher coefficient level than for the group with a low coefficient, that is, $t_L/\theta < t_H/(1-\theta)$. Inequality arises by adjusting the number of persons for the respective sizes of each category. The result obtained from this inequality is related to the average coefficient in each group, where $\bar{p}^L > \bar{p}^H$. For incompatible events $H_1, H_2, \ldots, H_i$ that constitute a complete group, a posteriori probability is calculated by the Bayes theorem, where $P(H_i)$ is the a priori probability of the event H_i, $P(E_j|H_i)$, $j=1,\ldots,J$ is the conditional probability of the evident E_j, provided that the event H_i occurred, and the event evident E_j has a nonzero probability (i.e. $P(E_j)>0$). Before applying the NBA, special attention should be given to preprocessing data. Initially, it is necessary to determine the a priori distribution of the desired multidimensional parameter; for a fixed parameter, to obtain the initial statistical data x_1, $x_2, \ldots, x_n$ with the corresponding distribution laws; to calculate the corresponding probabilities, create a likelihood table and using the Bayes theorem compute a posteriori probability for each data class. Work with data means "elimination" of unnecessary data by promotion of some hypotheses and their subsequent verification on the validity. In the considered case we denote these hypotheses as $\{H_t, t=1,\ldots,I\}$; the evidence supporting these hypotheses will be denoted as $\{E_j, j=1,\ldots J\}$. Each t-th hypothesis is confirmed by suggested hypotheses, a priori probabilities of the truth of the suggested hypothesis, a posteriori probabilities of the truth of the suggested hypothesis, the probability of realization of the evidence even in the case of hypothesis' disproof. The key element of the considered methodology is the calculation of the price of evidences $O(E_j)$, $j=1,\ldots,J$. The Bayesian approach is based on the idea of the validity of the suggested hypothesis. Even a negligible a priori probability can successfully transform into a posteriori probability when big data are received (Ramachandramurthy, Subramaniam, Ramasamy, 2015). The advantage of this approach is a consistent "enrichment" of the proof of hypotheses' validity, when a priori probability at each iteration of processing by big data, being guided by evidences, is modified into a posteriori. As an example, we study the impact of information asymmetry in the university on the future employment of young specialists. Our task is finding out which hypothesis with a particular evidence is reliable. The research is being conducted on data taken from (Trifonov, 2015). First, we need to find out the reasons for the appearance of asymmetry in the university. As reasons we accept the following hypotheses:

- H_1 – information asymmetry in the delivery of educational material;
- H_2 – information asymmetry in the perception of the educational material (the dynamics of the perception of the teacher);
- H_3 – static information asymmetry due to the state (the initial state of the student is characterized by the information asymmetry of professional self-determination).
- The following statements will act as evidences:

- E_1 – value of repeatability factor of the same transactions for decrease of information asymmetry in education market;
- E_2 – low characteristics of subjects' awareness;
- E_3 – overcoming "information gap" between labor market and education.
- E_4 – influence of computer literacy.

The initial a priori information about specific hypotheses, the evidence obtained after processing by naïve algorithm, and the subsequent calculations, we fix in the following table:

Table 1: The initial a priori information about specific hypotheses and evidences (*Authors*)

H	**1**			**2**			**3**		
P(H)	0,92	0,92	0,92	0,81	0,81	0,81	0,6	0,6	0,6
No. of evidences	0,8	0,15	0,1	0,8	0,1	0,6	0,15	0,1	0,6
P(E|H)	0,4	0,8	0,9	0,3	0,8	0,8	0,9	0,6	0,8
P(E|¯H)	0,8	0,15	0,1	0,6	0,12	0,4	0,4	0,2	0,4
P(¯E|H)	0,6	0,2	0,1	0,7	0,2	0,2	0,1	0,4	0,2
P(¯H|¯E)	0,2	0,85	0,9	0,4	0,88	0,6	0,6	0,8	0,6
P(H|E)	0,8	0,98	0,99	0,68	0,96	0,89	0,77	0,81	0,75
P(H|¯E)	0,97	0,73	0,56	0,88	0,49	0,58	0,2	0,42	0,33
P(H|E) -P(H|¯E)	0,1	0,25	0,42	0,20	0,47	0,30	0,57	0,38	0,41
P_{max}(H)	0,99			0,99			0,9		
P_{min}(H)	0,13			0,13			0,04		
Class of hypoth.	Probable			Probable			Probable		

The results of Bayesian classifier are reflected in the next table:

Table 2: The prices of evidences after first, second and third iterations for $O(E_j)$, j=1,2,3 (*Authors*)

The price of evidences	***O(E₁)***	***O(E₂)***	***O(E₃)***
	0,321111	0,185341	0,170165
	0,825227	0,690449	0
	1,292974	0	0
	0,724738	0,566734	0,532106

After first iteration, the third evidence has the highest estimation (figure1) and we pass to the next iteration. In the constructed diagrams, columns denote the probabilities of the hypotheses.

Figure following on the next page

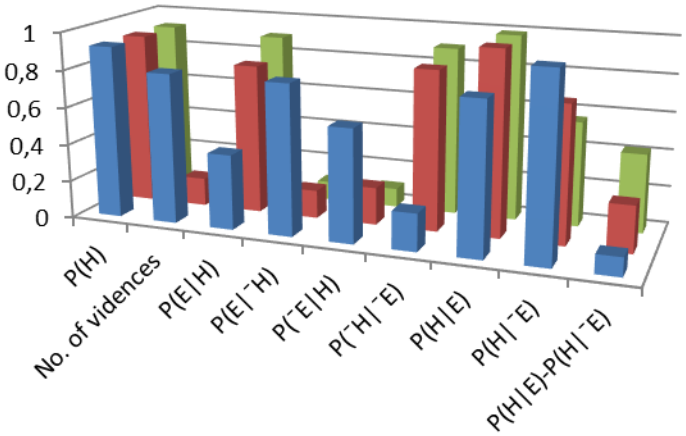

Figure1: The results of the first iteration (*Authors*)

After second iteration the second evidence has the highest estimate. We pass to the next iteration (figure 2).

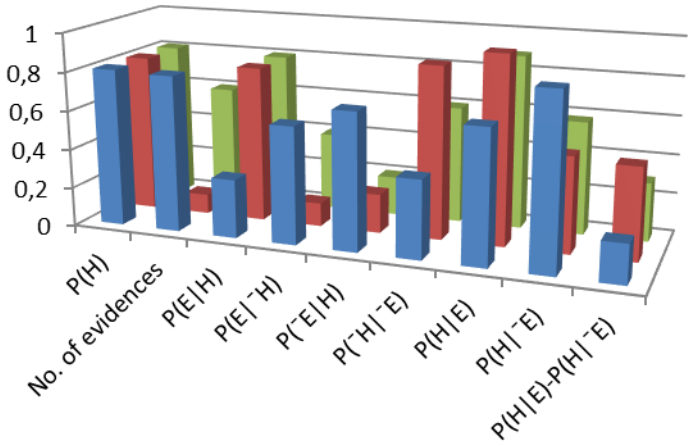

Figure 2: The results of the second iteration (*Authors*)

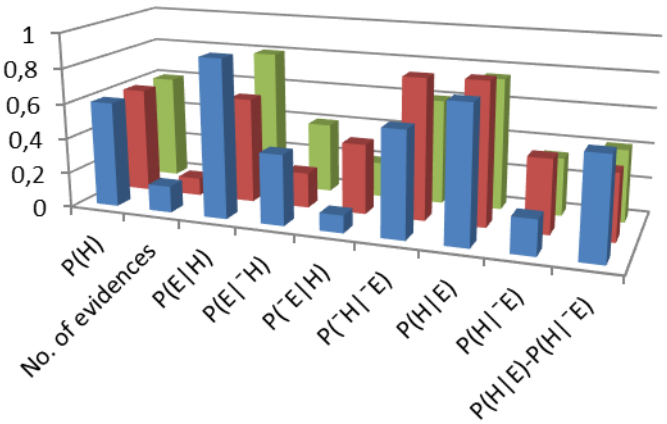

Figure 3: The results of the third iteration (*Authors*)

According to the calculations, the main reason for the information asymmetry is the static information asymmetry, and data visualization shows that the first hypothesis with the fourth evidence is reliable (figure 3).

5. CONCLUSION

In the very nature of the organization structure there is a certain "conflict of interests, reflected in the functional and professional differences between departments, the essence of individual, qualifying and business qualities of employees, the priority of personal interests over corporate ones, which can only be limited, but not excluded. On the asymmetric relations between the leaders and subordinates, the administrative vertical is built. Asymmetry as a characteristic feature of managerial relations promotes the emergence of social interests and forms a communicative mechanism of social progress. The modern environment is subject to constant changes and perturbations, which challenge traditional economic and business concepts. However, as the resources promoting distribution and acquisition of information increase there are more and more opportunities for decrease of its asymmetry. The article proposed a method for processing complex information queries using big data technologies based on the Bayesian approach, that make possible to extract the necessary information from many different sources, providing a higher return on their use, reducing the asymmetry of information and increasing its transparency. Bayesian approach may provide some results, even if there is absolutely no sample data. This is due to the use of the a priori probability distribution, which does not change in the absence of statistical data, and the a posteriori distribution recounted by the Bayes theorem coincides with the a priori distribution. In real economic systems, the development of non-standard, relevant, effective and, most importantly, interpreted knowledge becomes paramount in order to support the adoption of proper managerial decisions. Using this approach, carrying out additional experiments to refine the data states in the study of the various nature of decision-making problems is possible. The most important resource and tool for acquiring this knowledge is a deep and comprehensive analysis of data describing possible cause-effect relationships in social and economic systems. This is exactly the role that modern information technologies play today.

LITERATURE:

1. Akerlof G.A. (1970). The Market for «Lemons»: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics, Vol. 84 (3), pp. 488-460.*
2. Aliyeva T. (2015). Application Bayesian Approach in Tasks Decision-Making Support. *American Journal of Applied Mathematics and Statistics, Vol. 3(6), pp. 257-262*, DOI: 10.12691/ajams-3-6-7.
3. Daas P., Puts M. (2014). Big data as a source of statistical information. *The Surv. Stat., Vol. 69, pp. 22-31*.
4. Foreman J. (2016). *Many figures: Analysis of large data using Excel*. Moscow: Alpina Pablisher.
5. Holmstrom B., Milgrom P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization, Vol. 7, pp.24-52.*
6. Kudzh S. (2016). Risks of information asymmetry. *Perspectives of Science and Education, Vol. 6 (24), pp. 23-27.*
7. Liu T. (2015). Summary of Big Data and Macro Economy Analysis. *Foreign Theoretical Trends, № 1, pp. 57–66.*
8. Mayer-Shenberger V., Kuker K. (2014). *Big data. A revolution that will change how we live, work and think.* Moscow: Mann, Ivanov, Ferberm, ISBN: 987-5-91657-936-9.

9. Muller H.L., Stallard P.W. and Warren D. (1996). Multitasking and Multithreading on a Multiprocessor with Virtual Shared Memory, *Proceeding HPCA '96 Proceedings of the 2nd IEEE Symposium on High-Performance Computer Architecture, Washington, DC, USA, February 03 - 07.*
10. Nikitina K. (2011). Asymmetry of information in the investment market: content and forms of manifestation. *Bulletin of Economics, Law and Sociology, No. 1, pp. 61-65.*
11. Ramachandramurthy S., Subramaniam S. and Ramasamy Ch. (2015). Distilling Big Data: Refining Quality Information in the Era of Yottabytes. *Scientific World Journal, Published online Oct 1,* DOI: 10.1155/2015/453597.
12. Trifonov Yu. (1998). *The choice of effective solutions in the economy in conditions of uncertainty.* N. Novgorod, UNN.
13. Vasyutinskaya S. (2016). Information asymmetry in educational technologies. *Educational resources and technologies, Vol. 3 (15), pp. 14-20.*
14. Widenhofer G., Ytterstad E. (2017). *Asymmetric Information in Insurance. The Impact of Big Data on Low-SES Individuals.* Monograph, Norwegian School of Economics Bergen, December.
15. Zhou T. (2013). Big Data Version 1.0, Version 2.0 and Version 3.0: Business Revolution under Subversive Change. *People's Tribune, № 10, pp. 24–25.*