UDC 519.87

Vestnik SibGAU
Vol. 16, No. 1, P. 137–142

SOLVING THE PROBLEM OF CITY ECOLOGY FORECASTING WITH NEURO-EVOLUTIONARY ALGORITHMS

D. I. Khritonenko*, E. S. Semenkin, E. V. Sugak, E. N. Potilitsina

Siberian State Aerospace University named after academician M. F. Reshetnev
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660014, Russian Federation
*E-mail: hdmitry.91@mail.ru

In this paper the problem of a city ecological condition forecasting based on the chemical composition of the air is considered. The procedure of solving this problem with artificial neural networks, grown by evolutionary algorithms is described. Several modifications of evolutionary algorithms and ensemble approach for neural predictor design allowing the increase of prediction efficiency are presented. Also an ensemble method for neural predictor design is considered to increase the efficiency. The existing methods for design of intelligence information technologies ensembles have been considered. The comparison of their efficiency is presented in the paper for a set of test problems. A modified approach for artificial neural network ensembles is proposed, which is different from known before with combined application of existing schemes and methods for ensemble organization. In the problem description the problem of large amount of missing values in the dataset is highlighted. To solve this problem, a modified genetic programming method is applied. The usefulness of this method is shown for the problem solving. The testing shows the efficiency of the presented approach compared to basic and ensemble models. One of the applications of the developed algorithm is the time series prediction. Many technical systems contain a large amount of dynamic parameters, and tracking and predicting these parameters is an important problem. The rocket and space technology is no exception, so the described algorithm is a useful data analysis instrument for it. The developed approach can be used as a method for individual predictors' creation, and also as way for combining the existing ones. It is shown that this approach allows increasing the accuracy of the resulting models.

Keywords: artificial neural networks, ensembles, evolutionary algorithms, time series prediction, city ecology analysis.

Вестник СибГАУ
Т. 16, № 1. С. 137–142

РЕШЕНИЕ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ЭКОЛОГИЧЕСКОГО СОСТОЯНИЯ ГОРОДА НЕЙРОЭВОЛЮЦИОННЫМИ АЛГОРИТМАМИ

Д. И. Хритоненко*, Е. С. Семенкин, Е. В. Сугак, Е. Н. Потылицына

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева
Российская Федерация, 660014, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
*E-mail: hdmitry.91@mail.ru

Рассматривается решение задачи прогнозирования экологического состояния города в зависимости от химического состава воздуха. Описывается процедура решения данной задачи при помощи искусственных нейронных сетей, выращенных эволюционными алгоритмами. Представлены некоторые из модификаций эволюционных алгоритмов, которые позволяют повысить эффективность предсказания. Также с целью повышения эффективности рассматривается коллективный подход к построению нейросетевых предикторов. Были описаны существующие сегодня методы формирования коллективов интеллектуальных информационных технологий. Представлено сравнение их эффективности на ряде тестовых задач. Предлагается модифицированный подход к проектированию коллективов искусственных нейронных сетей, отличающийся от известных комбинированным использованием существующих схем и методов организации коллектива. В описании решаемой задачи подчеркивается проблема наличия большого количества пропусков в таблице данных. Для ее решения используется модифицированный метод генетического программирования. Показана полезность использования данной процедуры при решении поставленной задачи. Тестирование показывает эффективность предложенного подхода по сравнению с базовой и коллективной моделями. Одно из приложений разработанного алгоритма – предсказание временных рядов. Многие технические системы содержат большое количество динамических параметров, отслеживание и предсказание которых является важной задачей. Ракетно-космическая техника не является исключением, а потому описанный алгоритм является для нее полезным инструментом анализа данных. Разработанный подход может являться как средством создания самостоятельных

предикторов, так и средством комбинирования уже существующих. Показано, что такой подход позволяет увеличить точность итоговых моделей.

Ключевые слова: искусственные нейронные сети, коллективы, эволюционные алгоритмы, прогнозирование временных рядов, анализ экологической ситуации в городе.

Introduction. The modern ecological condition of cities leads to the necessity of finding the risk factors and developing a control system for solving the technogenic problems. However, today the methods for estimating the risk level for urban territories, including geo-ecological risks, are still not developed properly. The lack of generalizing representations and concepts and general methodological principles is one of the main disadvantages in the field of ecological risk estimations.

As most industrial and urban territories are located in basins, there may be increased local circulation and aggregation of impurities in lower parts of the relief [1]. Adverse orographic and meteorological features in connection with atmosphere air quality lead to a higher air pollution level in the following cases: industrial sites are close to residential areas; the number of cars is increasing; continuous house building and destruction of green areas. The high prevalence of diseases related to ecology among the populations of large industrial centres, and a trend of an increasing pathology level determine the special importance of studying the health of city populations.

Concerning this, there appears a real necessity of adding programmes for analysing information about the environment condition to the ecological monitoring complex. This necessity is due to the multi-dimensionality and multi-coupling of ecological data. Many ecological processes are characterized by non-linearity and uncertainty which makes estimation and forecasting the ecological situation more complex. To increase the efficiency and accuracy of the decision-making process, artificial intelligence technologies capable of working in the conditions of non-certain input information are used nowadays. One of the perspective directions of using artificial intelligence technologies in ecological monitoring is in the use of artificial neural networks in problems of predicting and forecasting ecological situations [2].

Automatic generation of artificial neural networks and their ensembles. An artificial neural network (ANN) is a combination of mathematical models of neurons of the human brain, which allow the most complex data analysis problems to be efficiently solved [3]. However, when using this approach the researcher always faces a problem of choosing the optimal structure of the neural network and the optimal learning procedure according to the specified efficiency criterion. Of course, the "trial and error" method is not a successful choice in this case, as it takes a huge amount of computational resources. The deep knowledge and experience of the expert in the field of neural networks is needed, nevertheless, the "trial and error approach" might result in insufficient modelling accuracy. Today for solving the given problem, methods imitating the evolutionary process of the real world – evolutionary algorithms – are used. They allow the researcher to design the artificial neural networks automatically and with high efficiency [4]. Also, today there is a set of modifications of evolutionary algorithms that allow their efficiency to be increased. Also, these algorithms do not require any parameter tuning, so they are easy to use for many researchers, even if they do not have specific knowledge in the field of evolutionary modelling [4; 5].

The standard scheme for solving data analysis problems [6] with intelligence information technologies (IIT), in particular, neural networks, involves two stages. In the first stage, the structure and parameters of the neural networks are adjusted, and in the second the best of the neural networks is used for solving the problem. However, using several neural networks at the same time by combining them into an ensemble can be useful [7; 8]. Increasing the quality of the received solutions with such an approach was observed long ago. The combining of neural networks into ensembles was first reported in the works of Hansen and Salamon [9]. The efficiency of such approaches was also shown in these works.

Today using ensembles is one of the promising directions in automated learning methods [10]. To design an ensemble, three tasks should be performed:

1) Choose the method of designing individual ANNs;

2) Selection of ANNs for forming ensembles from the overall set of solutions obtained;

3) Determining the method for consideration of the "views" for every part of such ensemble.

In 1988 the problem of weak learning capabilities was studied. The problem was solved by strengthening the weak algorithms with boosting [11]. The essence is that if this algorithm is in a specific method for forming the learning sample for every next neural network, the learning sample includes incorrectly solved examples from the previous one. So, the examples are processed by consecutive set of ANNs. In 1993 the efficient learning algorithm, based on this principle, was developed [12; 13]. In 1996 Breiman [14] proposed a bagging method, which learns every IIT with a sample, randomly (with returns) formed from the original one. This method is considered to be the most effective for samples of small size, as the lack of the part of learning examples in these samples leads to significant diversity in the received IIT. With a large number of features and a small number of learning examples the method of random subspaces should be used [15]. In such method separate IIT are learned on different sets of attributes, which were selected randomly. The listed methods allow the generation of different IITs (in particular, ANNs) for forming ensembles.

The question of how to consider separate views of different ensemble members is still to be considered. The basic work in this field describes the voting method [9], and also the averaging method. There is a modification of this method that allows errors of every ensemble member to be taken into account. We will now describe these two methods:

1. The outputs of the ensemble members are summed and divided by their number.

2. The error of each of the ensemble member is taken into account. Each of them is associated with a weight, inversely proportional to the error [16]. The output values are summed and divided by their number.

There is a scheme, allowing an error to be obtained which is not higher than the best of the ensemble members [7]:

1. The point in the learning sample is found which is the closest to the point where we need to get the value.

2. The IIT which has the lowest error is determined in this point.

3. The selected IIT makes the decision about the value in the point of interest.

The forming of the ANN ensemble is also possible by means of the self-configured genetic programming algorithm [4]. For this purpose, the functional set is filled with different mathematical operators and functions (sine, cosine, addition, subtraction, and so on), and the terminal set is filled with different adjustable coefficients and outputs of the neural networks which were selected into the ensemble. Using such an instrument does not require additional costs for its adjustment and allows ensemble forming in automatic mode.

Thus, applying evolutionary algorithms may allow the design of efficient ensembles of artificial neural networks in automated mode without using special knowledge of neuro-evolutionary modelling. The efficiency of the programme realization of the presented algorithms for ecological condition forecasting using the chemical air composition will be presented in the next section of the paper.

The problem description. There is a sample (x_i, y_i), $i = 1,2, \ldots, n$, where n – is the number of instances in the sample; x_i – is the vector of size m, where m – is the number of sample attributes, y_i – is the observed value. It is also known that y_i may not only depend on its previous k values, but also on values of x_i. The dependence between x_i, x_{i-1}, …, x_{i-k}, y_i, y_{i-1}, …, x_{i-k} and y_{i+1} should be defined, that means to predict the values of y_{i+1} or solve the time series problem [17].

The sample is a matrix with 39 columns, 28 of which are problem attributes, and 10 are the values to be predicted, and 32 rows – the number of instances. One of the columns is the year when the attributes were measured. There are also missing values in the database. The attributes show the constituents of different harmful materials in the air: copper, zinc, iron, chlorine, and so on. All other measures show different diseases and their mortality.

The data available in the sample was obtained in the environment pollution monitoring centre of GU "Krasnoyarsk CGSM-R" which observes the atmosphere air quality at 8 stationary posts in the city of Krasnoyarsk. The yearbooks of the "condition of the air pollution of the cities on the territory of Krasnoyarsk region, republics of Khakassia and Tuva" from year 1999 to 2012 were processed [18]. Sanitary-demographic statistics was obtained in the territorial authority of the Federal Service of State Statistics in Krasnoyarsk region ("Krasnoyarskstat") [19].

For the convenience of further understanding the data, the predicted values are numbered with values from *Out1* to *Out10* and some numerical data about these values is presented. In tab. 1 the original names of the predicted values are presented, as well as the corresponding numbering, the minimum and maximum values in the column and also the average value and the standard deviation.

Problem solving and results. The database for the problem has missing values, as already has been mentioned. Given the small size of the sample, and also the fact that the missing values are present in a significant part of the sample, the recovery of missing values is necessary. One of the promising approaches for the described conditions is the use of evolutionary algorithms [20], and in particular, the genetic programming algorithm for building non-linear regression. For this problem the terminal and the functional set of the genetic programming algorithm is filled with different mathematical operators and functions, and also some numerical values and feature variables of the problem. Here we provide the comparison of the prediction results by a perceptron, which was grown with a genetic programming algorithm [4] for the cases with and without the recovery of missing values. In the case without the use of the recovery procedure, the incomplete data was excluded, as well as the problem attributes.

Table 1

Numerical characteristics of problems

Predicted value	Birthrate (per 1000 of population)	Mortality (per 1000 of population)	Mortality for children under 1 (per 1000 of population)	Patients with malignancies (per 100.000 of population)	Mortality from malignancies (per 100.000 of population)
Numbering	Out1	Out2	Out3	Out4	Out5
Min.	8.3	8.9	4.8	1042	180.26
Max.	14.3	15.5	21.0	2483.8	211.9
Average	10.98	12.41	14.26	1822.65	202.44
Std. dev.	1.89	1.80	5.91	429.42	8.75
Predicted value	Mortality – circulatory system (per 100.000 of population)	Mortality – respiratory (per 100.000 of population)	Mentally ill (per 100.000 of population)	Mortality for mentally ill (per 100.000 of population)	Mortality – nervous system (per 100.000 of population)
Numbering	Out6	Out7	Out8	Out9	Out10
Min.	441.44	38.69	686.7	0.91	5.31
Max.	726.05	72.16	1407.4	11.42	8.98
Average	594.03	55.93	931.07	5.83	6.79
Std. dev.	76.16	8.70	198.96	3.91	1.13

A series of runs using the cross-validation procedure was performed for comparison for every value to be predicted. The average values of the following criterion over all runs are presented in the tab. 2:

$$I = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - y_i^*}{y_i^*} \right| \cdot 100\ \%,$$

where y_i – is the obtained value in the i-th point; y_i^* – is the true value; n – is the sample size, $y_i^* \neq 0 \forall i$. As we can see from the results presented in tab. 2, the missing values recovery procedure allows information about predicted values to be saved as well as the increased efficiency of the obtained predictors.

Now we will describe the proposed method for solving the problem and present the tested variants of the algorithm. The basic model is the procedure of generating automated artificial neural networks with arbitrary structure and different activation functions by means of genetic programming and a genetic algorithm. The tab. 3 below contains the average value of the required criterion over all runs for the best generated neural networks.

We will now describe the ensembles. For the dependence recovery procedure, the previously described methods for ensemble forming were tested (schemes 1–3). The comparison has shown that the modified genetic programming algorithm [4; 5; 8] handles this problem an order of magnitude better. This is why this method was chosen to build the neural predictors further (tab. 4).

It should be mentioned that it is difficult to track the collective diversity in such an approach. In part, this problem is solved by the ensemble forming algorithm itself. Also, the diversity is achieved by the combined use of GP and GA for the growing of neural networks. The stochastic nature of the evolutionary algorithms also helps to solve the problem of diversity.

To solve this problem, a more complex method of the ensemble organization is proposed. In tab. 3 it is called the "proposed method". As mentioned earlier, for maintaining the diversity in the case of small samples, the bagging method is promising [14]. For the predictor strengthening, the boosting approach can be used [11]. In this case, the ensemble organization scheme described in [7] is used. The final ensemble is formed with a genetic programming algorithm. The functional set in this case is a set of mathematical operators and functions. The terminal set contains coefficients and generated neural networks (or ensembles, made by boosting or bagging). Applying the boosting approach gives a minor, but statistically significant increase in efficiency (less than 1 %).

As we can see from tab. 3, the proposed method is 5 % better than the classical method for the ensemble design with GP on average. This method is, in turn, 13 % better than the basic model. The improvement is achieved because of the combined use of the bagging and boosting approaches. An important fact is that this approach adds only one additional parameter to the algorithm, which means that the optimization system programme does not become more complicated for the user.

Further improvement of the approach includes adding dynamics into the model by using recurrent neural networks. The flexibility of evolutionary algorithms allows such a kind of neural networks to be generated. Increasing the sample size will allow improvement of the existing predictors and the creation of new predictors for this problem.

Table 2

Missing values recovery algorithm efficiency

	Out1	Out2	Out3	Out4	Out5	Out6	Out7	Out8	Out9	Out10
With missing values	9.8	12.7	6.1	4.2	6.3	1.5	0.5	5.3	4.7	10.5
Without missing values	9.2	12.5	5.4	2.4	5.4	1.0	0.2	4.8	3.8	8.5

Table 3

Efficiency comparison

	Out1	Out2	Out3	Out4	Out5	Out6	Out7	Out8	Out9	Out10
Basic model	9.2	12.5	5.4	2.4	5.4	1.0	0.2	4.8	3.8	8.5
Ensembles	5.9	10.1	5.4	2.2	4.0	1.0	0.2	3.1	3.6	8.0
Proposed method	5.1	9.3	5.4	1.9	3.9	1.0	0.2	2.8	3.6	7.6

Table 4

Comparison of ensemble forming methods

	Scheme 1	Scheme 2	Scheme 3	GP
Mean square error	0.002	0.0015	0.004	0.0001

Conclusion. The real-world problem of the predicting the ecological condition of the city was solved with algorithms of neuro-evolutionary modeling. The standard methods of solving the problem were considered, as well as a modified method of ensembling artificial neural networks. The efficiency of the existing ensemble methods was tested. As a result of testing, for the most of the predicted values the proposed modification has shown better performance than the basic algorithm. For a set of predicted values, the efficiency of the approach did not change. This may be because of the small volume of the sample, and also because of the absence of key variables of the problem. The further expansion of the sample will allow solving this problem. Recovering the missed values with the genetic programming algorithm also allowed increasing the efficiency of the approach. Generating recurrent neural networks and their ensembles is one of possible ways for the further development of the approach. However, one should remember that using a universal procedure for evolutionary design of neural network predictors significantly increases the computational complexity that can be unacceptable in some cases.

Acknowledgment. The work is fulfilled with the support of the Russian Foundation of Fundamental Researches (project 14-06-00256).

Благодарности. Работа выполняется при поддержке Российского фонда фундаментальных исследований (проект 14-06-00256).

References

1. Taseiko O. V., Mikhailuta S. V., Zakharov U. V. [Modelling of local conditions of pollution dissipation among city buildings]. *LAP LAMBERT Academic Publishing GmbH & Co. KG, Saarbrucken*, 2011, 122 p. (In Russ.).

2. Potilitsina E. N., Lipinskiy L. V., Sugak E.V. [Using artificial neural networks for solving applied ecological problems]. *Sovremennye problemy nauki i obrazovaniya*, 2013, no. 4, p. 51 (In Russ.).

3. Kriesel D. A. Brief Introduction to Neural Networks. Available at: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf (accessed 01.9.2014).

4. Khritonenko D.I., Semenkin E. S. Distributed self-configuring evolutionary algorithms for artificial neural networks design. *Vestnik SibGAU*. 2013, no. 4 (50), p. 112–116.

5. Semenkin E. S., Semenkina M. E. [Application of genetic algorithm with modified uniform recombination operator for automated forming of intelligent information technologies]. *Vestnik SibGAU*. 2007, no. 3, p. 27–33 (In Russ.).

6. Lippmann, R. P. An introduction to computing with neural nets. *IEEE ASSP Mag*., Apr. 1987, p. 4–22.

7. Popov E. A., Semenkina M. E., Lipinskiy L. V., [Decision making by an ensemble of intelligent information technologies], *Vestnik SibGAU*, 2012, no. 5 (45), p. 95–99 (In Russ.).

8. Semenkin E. S., Shabalov A. A., Efimov S. N., [Automated design of intelligent information technologies ensembles with genetic programming]. *Vestnik SibGAU*. 2011, no. 3, p. 77–81 (In Russ.)

9. Hansen L.K., Salamon P. Neural network ensembles. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 1990, no. 12 (10), p. 993–1001.

10. Zhou Z.-H., Jiang Y. NeC4.5: Neural Ensemble Based C4.5. *IEEE Trans. Knowl. Data Eng.* 16(6): 770–773, – 2004.

11. Schapire R. E. The strength of weak learnability. *Machine Learning*. 1990. Vol. 5. pp. 197–227.

12. Schapire R. The boosting approach to machine learning: An overview. *MSRI Workshop on Nonlinear Estimation and Classification, Berkeley, CA*. 2001.

13. Freund Y., Schapire R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *European Conference on Computational Learning Theory*. 1995. pp. 23–37.

14. Breiman L. Bagging predictors. *Machine Learning*. 1996, vol. 24 (2), p. 123–140.

15. Skurichina M., Duin R. P. W. Limited bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications*. 2002, no. 5, p. 121–135.

16. Perrone M. P., Cooper L. N. When networks disagree: ensemble method for neural networks. *Artificial Neural Networks for Speech and Vision, Chapman & Hall, New York,* 1993, p. 126–142.

17. Semenkin E. S., Shabalov A. A. Intelligent information technologies in time series forecasting. *Vestnik SibGAU*. 2013, no. 4 (50), p. 128–134.

18. *Sostoyanie zagryazneniya atmosfernogo vozdukha gorodov na territorii Krasnoyarskogo kraya, respublik Khakasiya i Tyva v 2010 g.* [The condition of air pollution of the cities on the territory of Krasnoyarsk region, republics of Khakassia, Tuva in 2010]. Krasnoyarsk, 2011, 116 p.

19. *Zdravookhranenie v g. Krasnoyarske v 2012 godu. Territorial'nyy organ federal'noy sluzhby gosudarstvennoy statistiki po Krasnoyarskomu krayu* [Health service in Krasnoyarsk in the year 2012. Territorial authority of federal service of state statistics in Krasnoyarsk region]. Statistic Bulletin no. 8-5.2. Krasnoyarsk, 2013, 15 p.

20. Snituk V. E. [Evolutionary method for missing data recovering]. *Sbornik trudov VI-y Mezhd. konf. "Intellektual'nyy analiz informatsii"* [Proceedings of the VI-th International conference "Intelligent data analysis"]. Kiev, 2006, p. 262–271 (in Russ.).

Библиографические ссылки

1. Тасейко О. В., Михайлюта С. В., Захаров Ю. В. Моделирование локальных условий рассеивания загрязнителей в городской застройке // LAP LAMBERT Academic Publishing GmbH & Co. KG. Saarbrucken, 2011. 122 с.

2. Потылицына Е. Н., Липинский Л. В., Сугак Е. В. Использование искусственных нейронных сетей для решения прикладных экологических задач // Современные проблемы науки и образования. 2013. № 4. С. 51.

3. Kriesel D. A. Brief Introduction to Neural Networks [Электронный ресурс]. URL: http://www. dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf (дата обращения: 01.9.2014).

4. Khritonenko D. I., Semenkin E. S. Distributed self-configuring evolutionary algorithms for artificial neural networks design // Вестник СибГАУ. 2013. № 4 (50). С. 112–116.

5. Семенкин Е. С., Семенкина М. Е. Применение генетического алгоритма с модифицированным оператором равномерной рекомбинации при автоматизированном формировании интеллектуальных информационных технологий // Вестник СибГАУ. 2007. № 3. С. 27–33.

6. Lippmann R. P. An introduction to computing with neural nets // IEEE ASSP Mag. Apr. 1987. Pp. 4–22.

7. Попов Е. А., Семенкина М. Е., Липинский Л. В. Принятие решений коллективом интеллектуальных информационных технологий // Вестник СибГАУ. 2012. № 5 (45). С. 95–99.

8. Семенкин Е. С., Шабалов А. А., Ефимов С. Н. Автоматизированное проектирование коллективов интеллектуальных информационных технологий методом генетического программирования // Вестник СибГАУ. 2011. № 3. С. 77–81.

9. Hansen L. K., Salamon P. Neural network ensembles // IEEE Trans. Pattern Analysis and Machine Intelligence. 1990. № 12 (10). Pp. 993–1001.

10. Zhou Z.-H., Jiang Y. NeC4.5: Neural Ensemble Based C4.5 // IEEE Trans. Knowl. Data Eng. 2004. № 16(6). P. 770–773.

11. Schapire R. E. The strength of weak learnability // Machine Learning. 1990. Vol. 5. Pp. 197–227.

12. Schapire R. The boosting approach to machine learning: An overview // MSRI Workshop on Nonlinear Estimation and Classification. Berkeley, CA, 2001.

13. Freund Y., Schapire R. E. A decision-theoretic generalization of on-line learning and an application to boosting // European Conference on Computational Learning Theory. 1995. Pp. 23–37.

14. Breiman L. Bagging predictors // Machine Learning. 1996. Vol. 24 (2). Pp. 123–140.

15. Skurichina M., Duin R. P. W. Limited bagging, boosting and the random subspace method for linear classifiers // Pattern Analysis & Applications. 2002. № 5. Pp. 121–135.

16. Perrone M. P., Cooper L. N. When networks disagree: ensemble method for neural networks // Artificial Neural Networks for Speech and Vision, Chapman & Hall. New York, 1993. Pp. 126–142.

17. Semenkin E. S., Shabalov A. A. Intelligent information technologies in time series forecasting // Вестник СибГАУ. 2013. № 4 (50). С. 128–134.

18. Состояние загрязнения атмосферного воздуха городов на территории Красноярского края, Республик Хакасия и Тыва в 2010 г. : ежегодник. Красноярск, 2011. 116 с.

19. Здравоохранение в г. Красноярске в 2012 году : стат. бюллетень № 8-5.2 / Территориальный орган Федеральной службы государственной статистики по Красноярскому краю. Красноярск. 2013. 15 с.

20. Снитюк В. Е. Эволюционный метод восстановления пропусков в данных // Интеллектуальный анализ информации : сб. тр. VI Междунар. конф. Киев, 2006. С. 262–271.