Machine learning assessment

Vladimir Konyukhov[1], Daria Musatova[2], Anna Zueva[3], Alexey Sorokin[4] and Serik Toygambayev[5]

[1] Russian State University of Physical Education, Sport, Youth and Tourism, 4, Lilac Boulevard house, Moscow, 105122, Russian Federation
[2] Lomonosov Moscow State University, 1, Leninskie Gory, Moscow, 119991, Russia
[3] Financial University under the Government of the Russian Federation, 49, Leningradsky avenue, Moscow, 125993, Russian Federation
[4] Academy of Civil Protection of the Ministry of Emergency Situations of Russia, 1, Sokolovskaya, Khimki, 141435, Russian Federation
[5] Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Larch Alley str., 16A, bldg. 3, sq. 409, Moscow, 127550, Russian Federation
nk-kfea@mail.ru

Abstract. This article discusses the concept of machine learning, its main characteristics and types, as well as the application of machine learning to artificial intelligence units. The research raises an important question about the legal regulation of providing access to data in the machine learning of artificial intelligence units.

Keywords: machine learning, data access, artificial intelligence, data, personal data, algorithm, legal regulation.

1   Introduction

Machine learning nowadays plays a very important role in people lives and definitely is going to take a leading role in the future. Specialists in the development of learning algorithms are already considered among the most in-demand professions. Information is a basis for data and due to the fact that there is a lot of it makes hard for people to over think and study it using only their mind. But how does it work? Firstly, machine learning optimize data by selecting information and putting it in order using a spectrum or range. This makes it more convenient for a person because is decreases an amount of time needed to do it manually, so machine learning technologies are in great demand today. Machine learning technologies are being actively implemented in such important areas for society as medicine, transport, education, and agriculture. In practice, this process is hampered by objective problems, many of which lie in the legal field. The article discusses some of them [1-6].

---

2 Materials and methods

To consider the concept of machine learning and its classifications, as well as to find a way to regulate the access to data used in machine learning of artificial intelligence units in the legal sphere, the author uses such methods as general scientific (deduction), special (structural and functional) and private legal (formal legal, comparative legal) methods.

This problem is studied by a lot of legal scholars McCarthy J., Robertson J., Mitcheland T. others.

3 Results

The term "Machine learning" stands for a data analysis method that helps to automates the process of creating an analytical model. This autoimmunization improves with time as artificial intelligence is made up in such way that it can learn and adapt through experience. This process of collecting information and statistics assembled by artificial intelligence with the help of which it can make predictions of any type is usually messed up with the term "data mining" [7]. But there is a difference that consists in the fact that data searching is used in cases where we need to discover unknown or hidden elements and to sum up this information and the information that is brought in data mining is specifically extracted for people and in machine learning artificial intelligence uses this information for improving it own processes [8]. Creating and improving an algorithms of machine learning is the main essence of presenting new data [2].

Machine learning methods include four categories [6]:

— Supervised machine learning algorithms. They act on the principle that they apply previously analyzed information to new data, using labeled examples to predict future events. The machine learning algorithm creates an inferred function to predict the output values, starting with the analysis of a known training data set. The system is able to provide targets for any new input after sufficient preparation. The learning algorithm can also compare its output signal with the correct, assumed output signal and find errors to modify the model accordingly.
— In contrast, when the information used for training is not classified or labeled, unsupervised machine learning algorithms are used. Unsupervised learning studies how the system derives functions describing hidden structures from unlabeled data. The system will not find the correct results, but will examine the data and draw conclusions from the data records in order to describe the hidden structure in the unlabeled data.
— Reinforcement machine learning algorithms is a type of a learning method of artificial intelligence that produces actions and finds errors or rewards. The software interacts with a modern day world and its changes due to this it has to deal with a lot of specific tasks. The example of this can be a self-driving car and its systems that controls the distance to the next car. The better the process of preventing errors in the work the correctly the process of reinforcement machine learning is and

makes artificial intelligence determine the ideal behavior. The choice of agent to find out which action is the best, simple reward feedback is needed-this is called a reinforcement signal.

— Semi-supervised machine learning algorithm is balance between supervised and unsupervised learning, because they use both labeled and unlabeled data for training – usually a small amount of labeled data and a large amount of unlabeled data. This greatly improves the accuracy of training. When the generated labeled data requires qualified and appropriate training/learning resources, semi-supervised learning is usually chosen. Otherwise, obtaining unlabeled data usually does not require additional resources [9-15].

Machine learning is at the base of the decision-making system of artificial intelligence algorithms. Learning is present at different stages of the AI life cycle, and the subsequent accuracy of its work depends on how the learning process is built. The data on which the algorithm is trained forms its experience, which means a direct link between the data for training and what decisions will eventually be made by the algorithm [4]. A variety of AI systems and robots can only learn effectively if they are provided with a very large amount of data for appropriate processing. The more data is loaded into the algorithm, the more effective the learning process will be. Today, there is a steady trend in the world towards the legislative allocation of an increasing amount of data, access to which is restricted: personal data, medical data, geolocation and information containing the secret of communication, other information that can primarily serve as identification of subjects and objects [5]. Often, it is the use of such data that is associated with the main breakthroughs in the field of machine learning. In this regard, the question arises: should access to data that is necessary for training and for solving problems in socially important areas be provided on special, simplified grounds?

4 Discussion

According to a number of scientists, to solve the problem, it is necessary to develop special rules that would define the limits of the use of AI data and allow us to maintain a balance between the principle of preserving privacy and the availability of a variety of human data necessary for the development of machine learning technology [1; 12]. The data can be protected by means of "safe harbour", the need to obtain the consent of the data subject, as well as the requirements for minimizing the use of data. Also, one of the ways to solve this problem is the depersonalization of personal data. However, depersonalization of personal data does not always guarantee their full protection. There are a number of cases when comparing depersonalized data with each other, it was possible to reveal the subjects of such data [16-20].

An important factor in this area is also the fact that large corporations have much more data processing capabilities, they have access to more data and are automatically in the most advantageous position, which indicates monopolization. This leads to the creation of a dominant position for a small number of large companies that use Both to collect information about their users and in their actions are often guided only by their own interests and internal regulations [3].

To solve the problem of artificial intelligence bias, it is necessary to ensure that the data provided for training is as objective as possible [9]. And it should not be trained on any available information: it is necessary to take a responsible approach to the selection of data and choose only those that exclude subjective assessments as much as possible. In this regard, it is necessary to develop standards for the data on which AI systems will be trained. The data used must be checked for compliance with this standard, which must meet the requirements of applicable EU law, including the GDPR [21-24].

The machine learning process should be organized only on the basis of reliable data and only using scientifically-based and proven algorithms, while the amount of data for training should be obviously sufficient, as well as a ban on illegal interference in the learning processes should be established and a complete and reliable recording of all information collected and processed by artificial intelligence, and information about the choice and adoption of all its decisions in the learning process should be organized.

5   Conclusion

The questions that arise in connection with machine learning make us think about a global problem: already there is a clear conflict between the need to respect basic human rights – the right to privacy, restriction of access to personal data, on the other hand, the need in some special situations to sacrifice such rights for the sake of scientific progress and the development of society as a whole. Of course, in this conflict, the key task is to find a balance of interests: it is necessary to find a "point of balance" between these two vectors of technology development.

In any case, the training of algorithms should be based on ensuring respect for fundamental human rights.

References

1. Asaro, P.: Robots and Responsibility. Legal Perspective, 11-14 (2007).
2. Dorschel A.: Rethinking Data Privacy: The Impact of Machine Learning. Luminovo A. I., 211-214 (2019).
3. Kingston, J.K.C.: Artificial Intelligence and Legal Liability. Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems XXIV, 270–280 (2016).
4. Krensky, P., Hare, J.: Hype Cycle for Data Science and Machine Learning, 47-52 Gartne (2018).
5. Solum, L.: Legal Personhood for Artificial Intelligences. North Carolina Law Review, 1231-1287 (April 1992).
6. Lipton, Z., Steinhardt, J.: Troubling trends in machine learning scholarship. ICML, 110-114 (2018).
7. Mitchell, T.: Machine Learning. A Guide to Current Research. Tom M. Mitchell, Jaime G. Carbonell, Ryszard S. Michalski (Eds.). Springer Science & Business Media, 178-182 (1986).

8. Robertson, J.: Human Rights vs. Robot Rights: Forecasts from Japan. Critical Asian Studies, 571–598 (2014).

9. Willick Marshall S. Artificial Intelligence: Some Legal Approaches and Implications, 134-142 (1983).

10. McCarthy, J.: Machine Learning. What it is and why it matters, https://www.sas.com/en_us/insights/ analytics/machine-learning.html, last accessed 2021/06/20.

11. Zhichkin, K., Nosov, V., Zhichkina, L.: The production costs calculation automation for planning the crops production parameters. CEUR Workshop Proceedings 2843, 20 (2021).

12. McCarthy, J. What is artificial intelligence? http://www-formal.stanford.edu/jmc/whatisai/, last accessed 2021/06/20.

13. Sadriddinov, M.I., Mezina, T.V., Morkovkin, D.E., Romanova, Ju.A., Gibadullin, A.A.: Assessment of technological development and economic sustainability of domestic industry in modern conditions. IOP Conference Series: Materials Science and Engineering, 734, 012051 (2020).

14. Fokicheva, A., Abramov, V., Istomin, E., Sokolov, A., Goloskvskaya, E., Levina, A.: Machine learning with digital generators for training sets including proteins modeling in the context of big data and blockchain technologies. Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision, 2020, 8638-8642 (2019).

15. Zhichkin, K., Nosov, V., Zhichkina, L., Panchenko, V., Zueva, E., Vorob'eva, D.: Modelling of state support for biodiesel production. E3S Web of Conferences 203, 05022 (2020).

16. Ermakova, A., Oznobihina, L., Avilova, T.: Analysis of the current state and features of natural resource potential management. E3S Web of Conferences, 157, 3005 (2020).

17. Khayrzoda, S., Morkovkin, D., Gibadullin, A., Elina, O., Kolchina, E.: Assessment of the innovative development of agriculture in Russia. E3S Web of Conferences 176, 05007 (2020).

18. Zhichkin, K., Nosov, V., Zhichkina, L., Pavlyukova, A., Korobova, L.: Modeling the production activity of personal subsidiary plots in the regional food security system. IOP Conference Series: Earth and Environmental Science 659, 012005 (2021).

19. Istomin, E.P., Burlov, V.G., Abramov, V.M., Sokolov, A.G., Bidenko, S.I.: Decision support model within environmental economics. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM 19(5.3), 139-145 (2019).

20. Zimnukhova, D.I., Zubkova, G.A., Morkovkin, D.E., Stroev, P.V., Gibadullin, A.A.: Management and development of digital technologies in the electric power industry of Russia. Journal of Physics: Conference Series 1399, 033097 (2019).

21. Istomin, E.P., Abramov, V.M., Lepeshkin, O.M., Baikov, E.A., Bidenko, S.I.: Web-based tools for natural risk management while large environmental projects. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, 19, 953-960 (2019).

22. Morkovkin D., Lopatkin D., Sadriddinov M., Shushunova T, Gibadullin A., Golikova O. Assessment of innovation activity in the countries of the world // E3S Web of Conferences. № 157. Pp. 04015 (2020).

23. An, D., Song, Y.,Carr, M.: A comparison of two models of creativity: Divergent thinking and creative expert performance. Personality and Individual Differences, 90, 78-84 (2016).

24. Popova, A., Abramov, V., Popov, N., Istomin, E., Sokolov, A., Levina, A.: Blockchain and big data technologies within geo-information support for arctic projects. Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision, 2020 2019, 8575-8579 (2019).