Effective scientific personnel training in the field of modern computer technologies for the implementation of advanced research projects of the Megascience class

A N Shtanko, S D Kulik and A A Kondakov
National Research Nuclear University MEPHI, Moscow, 115409, Russian Federation

shtanko-mephi@yandex.ru

Abstract. Successful research projects of the Megascience class usually require a well-trained team of scientists from various fields of knowledge. These scientists must be high-skilled experts. Each member of a team like that must have the necessary, specialized cross-industry skills, for example, in such areas as artificial intelligence, convolutional neural networks, specialized intelligent search engines, and full-text analysis. One of the key aspects of effective personnel training for successful implementation of Megascience projects into reality is the acquisition by students professional skills, abilities, and knowledge to use tools of modern scientific technologies, containing, for example, libraries of programs (functions). In particular, convolutional neural networks and intelligent search systems can be applied in various research projects in the field of physics, chemistry, biology, and medicine, for example, in telemedicine, for effective decision-making in diagnosing a patient. Therefore, understanding the principles of neural networks and intelligent search systems is a necessary competence of researchers working in the framework of Megascience projects. Classic search engines are based on indexing the textual information of the database that is being searched. Intelligent search engines can improve the search experience through intelligent data processing, including using convolutional neural networks. This report examines practical examples and areas of the successful application of convolutional neural networks and information systems in practice.

1. Introduction
Research and innovation in the modern world sometimes demand large numbers of scientists working together in order to achieve a common goal. Consequently, we also need additional staff to organize the operation of these projects and appropriately large funding. To organize, manage, and finance these projects cooperation between multiple large corporations or countries is required.

These big projects are called Megascience projects or projects of the Megascience class. Typical features of the Megascience projects are:

- Complex physical and scientific research facilities. The complexity of the goals realizes into a necessity for respectively complex equipment and facilities. Often the field of study is related to physics, for example, particle physics.
- These projects require significant financial investments. This requirement comes from equipment and staff costs.
Participation of multiple countries. The necessary funding and the nature of studies often mean that only governments are willing to participate.

Large amount of data. Reaching the goal requires a lot of scientific experiments with various equipment that generates data that needs to be organized, stored, and analyzed. Additionally, because of the size of the projects, there’s a lot of administrative data.

The need for new efficient algorithms for intelligent data processing based on, for example, neural networks.

To achieve the goals of Megascience projects effective computer science algorithms are necessary to analyze the generated data and maintain information systems of the management structure. Additionally, with the progress made in the field of artificial intelligence, some of the future Megascience projects might be entirely focused on artificial intelligence. Thus, new and effective computer science algorithms are useful for Megascience projects:

- Radio signal detection [1].
- Multilayer perceptron network [2].
- Memristor and artificial neural networks [3].
- Neural networks and convolutional neural networks [4, 5].
- Special generator of articles [6] and recognition for biological and criminalistics objects [7].
- Recognition for dynamic visual scene analysis [8].
- Modelling the calcination of phosphorite pellets in a dense bed [9].
- Data management system [10].
- Effective agent-based technologies [11].
- Special agent technology [12] or model of communication [13].
- Emotionally-Intelligent assistant [14].
- Cognitive constructor [15].
- Cloud System [16].
- Telemedicine [17] and robot-doctor [18].

Because of the scale, the large projects like these generate a lot of experimental and other additional data. We need an automated way to process this data because manual data processing is too time-consuming. Thus, we need efficient data processing algorithms. Personnel training is required for the development and use of these complex special algorithms. We need different training for the development of new algorithms and systems, the adaptation of already existing ones, and operating them.

Neural networks have great potential for various applications. Neural networks and artificial intelligence in general are rapidly developing areas, and have already proven their effectiveness in many areas, especially when intelligent data processing is involved. However, an important feature of neural network technologies is high requirements for computation capabilities of the hardware, both for neural network training and application. Neural network training can take hours and in some cases even months. Fortunately, more powerful hardware is becoming available for lower costs, new optimized algorithms allow for relatively better performance on weaker hardware.

Neural networks [4] were successfully applied in different fields, for example:

- Weather prediction, for example, rain intensity [19].
- Detection of anti-tank and anti-personnel mines on radar data: applying a typical convolutional neural network to radar images to detect underground military mines, taking advantage of the similarity between radar data and natural images [20].
- Fingerprint spoofing detection: differentiating between alive fingerprint presented to a scanning device and an imitation [21].
Artificial intelligence for board games: using convolutional neural networks and reinforcement learning to reach superhuman playing level (without external human knowledge) in traditional board games [22].

It’s worth noting that many practical problems can be solved with the neural network algorithm of the same architecture. For example, convolutional neural networks are designed to work with image-like data, and if we have two-dimensional data similar to images, it’s possible to apply algorithms developed to analyze natural images.

In the area of neural networks, we can observe the following tendencies:

• Transition from algorithms based on knowledge of the subject area to universal algorithms. As the ability of neural networks to generalize their knowledge increases, it becomes possible for one architecture to cover more potential application tasks.
• More powerful computing hardware is becoming more widely available. It allows us to run more computation-intensive, deeper architectures.
• New, more efficient algorithms appear regularly for various applications. More available architectures allow us to choose the best option for the specific task at hand.
• There is a well-established toolkit for the practical application of convolutional neural networks. New tools provide an opportunity to use neural networks in more application fields.

Since the traditional data storage was replaced with cloud services, the volume and availability of data have dramatically increased. Many services now keep a personalized history for each user, many applications store request history without time limits. These are good data for training neural networks. Now the development of the Internet and personalized applications makes it possible to study a specific user clearly and qualitatively and choose the necessary solution for them.

Since computer science and artificial intelligence are becoming an integral part of scientific research, we should develop training programs for scientists.

2. Personnel training

Since the purposes of the training can vary depending on the position and area of the expertise of the students, it makes sense to outline several different training levels:

• User of machine learning applications.
• User of machine learning tools.
• Application developer (using libraries).
• Machine learning algorithms developer.

The higher, shallow levels generally don’t need to know the details of the lower levels. Depending on the background and given task, the researchers should be taught fitting level.

Machine learning algorithms developer is the deepest, 4th level. This level requires deep knowledge of hardware and software, and it’s needed when developing new algorithms from the ground up. The scope of this training covers:

• Development of libraries, low-level optimization of algorithms.
• Mathematical background required.
• An understanding of the features of the equipment and hardware is required.
• It is in demand when creating fundamentally new machine learning algorithms.

On this level, for example, Theano library can be used. Theano has the following features:
• It’s a low-level library for numerical computation.
• Symbolic calculations.
• Efficient parallel computing on CPU / GPU.
• Mathematical operations optimization.
• Optimization for multidimensional arrays.
• A library for algorithms that require significant computing resources, for example, for neural networks.

The application developer is the next, 3rd level. On this level, the developers create tools and applications using libraries abstracting themselves from low-level programming. It has the following features:

• Development of applications and scripts (machine learning algorithms) using library functions.
• The developer doesn’t need to completely understand underlying math and other low-level features.
• The purpose of the algorithms must be precisely defined, or the results may be unexpected.

Tools like lasagne library can be used here. They provide high-level functions for the training and application of neural networks. Lasagne library is a relatively simple package for creating and training neural networks, it’s based on Theano. It provides the most common neuron layers (convolutional, fully connected ad so on), different learning methods, and other elements.

Another tool of the same level is RC4 (.NET) library for data encryption. It is an open-source data encryption library. It allows using synchronous encryption to transfer data in encrypted form between the client application and server application, or to encrypt the data before it is put into the database. The library is often used in client-server applications that work with personal data of users, for example, in applications for medical institutions [17]. Other advantages of this library are:

• Easily integrates into an existing client-server application to protect transmitted and stored data.
• Allows you to store data in the database in encrypted form.
• Provides fast interaction of system components.
• Allows you to protect the necessary elements of a Megascience class project.

Entity Framework is the most popular and flexible library from Microsoft for easy access to data from an application. Developers all over the world use it in their work. It allows the developer to build multi-tier applications and provides the developer with the most convenient design pattern. Entity framework has the following features:

• Support for all modern databases allowing integration with them for .Net application.
• Flexible settings with the ability to migrate to other subsections of the Megascience project.
• Support of modern technologies (dependency injection, automatic generation of classes based on DBFirst/CodeFirst entities, etc.).

DeepMorphy analyzer. Applications that work with customer’s requests must process natural language input. This is an extremely difficult and non-trivial task. Working with natural language requests is impossible sometimes, so such requests need to be converted. To facilitate such applications, you can use one of the text analysis libraries. DeepMorphy analyzer is one of them. It allows developers to process texts, bring words to normal forms; it also has a large dictionary and uses a neural network for analysis. This library will save a lot of time for development and help to avoid the
creation of own text handlers. This is a very convenient and useful library for specialists working in areas related to text analysis and processing user queries in natural languages. Main features of this library:

- Morphological parsing of words, form normalization.
- Supports the Russian language.
- Complex text arrays and single words processing.
- The core element of DeepMorphy is a neural network.
- For most words, morphological analysis and lemmatization are performed by a neural network.

A user of machine learning tools is trained to use the command-line and GUI tools to perform machine learning and other tasks without significant programming expertise:

- Ready to use algorithms and machine learning tools.
- Training and using complex neural networks without programming expertise.
- Computational resources are required for the training algorithm (neural network).

For example, an object detection system YOLO [23] provides a command-line tool that allows a user to train and test the object detection system for custom objects (without programming knowledge). YOLO has the following features:

- It has a command-line tool to control its features.
- Little programming knowledge is required to use.
- The user prepares the training dataset themselves.
- The user can choose different configurations of the neural network.

The user of machine learning applications. In this category users are taught how to use a specific machine learning application:

- Training to use a specific application for solving specific scientific problems.
- Teaching application’s specifics required to successfully perform scientific tasks.
- Teaching about “unexpected” errors of neural networks.
- Avoiding excessive trust in the algorithm.

3. Conclusion
We’ve shown that the interest in artificial intelligence and neural network technologies is growing. Convolutional neural networks are a versatile tool for solving a wide variety of complex practical problems. For many applications, convolutional neural network tools have already been developed that can be used as a black box. To train modern specialists, it is necessary to improve the learning process in the field of neural networks. We’ve identified four levels of personnel training for Megascience class projects. Examples of tools are provided for each level.

It is also very important to use previous experience when developing modern applications. This is especially true for applications related to text data processing and recognition. In modern conditions, a lot of personal user data is stored on the Internet. When working with them, the developer needs to be extremely careful, therefore, to ensure the security of data transfer, they should use only proven open-source libraries. This will help avoid many problems and ensure the proper level of data protection. It is imperative for a specialist working in such areas to study the encryption, storage, and transmission libraries.
Acknowledgments
This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).

References
[1] Danilin S N, Shchanikov S A, Iventev A A and Zuev A D 2018 Advanced techniques of artificial networks design for radio signal detection Journal of Physics: Conference Series 1015 032168
[2] Danilin S N, Shchanikov S A, Zuev A D, Bordanov I A, Korolev D S, Belov A I, Pimashkin A S, Mikhaylov A N and Kazantsev V B 2019 Design of multilayer perceptron network based on metal-oxide memristive devices 12th International Conference on Developments in eSystems Engineering (DeSE) IEEE pp 533-8
[3] Danilin S N, Shchanikov S A, Bordanov I A and Zuev A D 2019 Quantitative determination of fault tolerance of memristor-based artificial neural networks Journal of Physics: Conference Series 1333 062027
[4] Shtanko A N and Kulik S D 2019 Scientific personnel training in convolutional neural networks for the implementation of research projects of the MegaScience class IOP Conference Series: Journal of Physics: Conference Series 1406 012014
[5] Trofimov A G and Bogatyreva A A 2020 A method of choosing a pre-trained convolutional neural network for transfer learning in image classification problems Studies in Computational Intelligence 856 pp 263-70
[6] Kulik S D, Belov A N and Matveeva K I 2018 Development of generation special short articles for the given topic International Journal of Engineering & Technology 7 2.23 (Special Issue 23) pp 171-4
[7] Kulik S D and Shtanko A N 2020 Recognition algorithm for biological and criminalistics objects Biologically Inspired Cognitive Architectures Proceedings of the Tenth Annual Meeting of the BICA Society, AISC 948 pp 283–94
[8] Borisov V V and Garanin O I 2018 A method of dynamic visual scene analysis based on convolutional neural network Proceedings of 16th Russian Conference on Artificial Intelligence, RCAI-2018 (Moscow, Russia) Springer: Communications in Computer and Information Science 934 pp 60–9
[9] Bobkov V I, Borisov V V, Dli M I and Meshalkin V P 2015 Modeling the calcination of phosphorite pellets in a dense bed Theoretical Foundations of Chemical Engineering 49 2 pp 176-82
[10] Artamonov A, Ionkina K, Tretyakov E and Timofeev A 2018 Electronic document processing operating map development for the implementation of the data management system in a scientific organization Procedia Computer Science 145 pp 248-53
[11] Ananieva A, Onykii B, Artamonov A, Ionkina K, Galin I and Kshnyakov D 2016 Thematic thesauruses in agent technologies for scientific and technical information search Procedia Computer Science 88 pp 493-8
[12] Artamonov A, Onykii B, Ananieva A, Ionkina K, Kshnyakov D, Danilova V and Korotkov M 2016 Regular agent technologies for the formation of dynamic profile Procedia Computer Science 88 pp 482-6
[13] Antonova L V, Klyuchnikov M M, Loktionov A A and Samsonovich A V 2018 Model of electronic objects in an intelligent virtual the-flag paradigm Procedia Computer Science 145 pp 72-6
[14] Krylov D I and Samsonovich A V 2019 Designing an emotionally-intelligent assistant of a virtual dance creator Advances in Intelligent Systems and Computing 848 pp 197-202
[15] Samsonovich A V, De Jong K A, Kitsantas A, Peters E E, Dabbagh N and Kalbfleisch M L 2008 Cognitive constructor: an intelligent tutoring system based on a biologically inspired
cognitive architecture (BICA) Frontiers in Artificial Intelligence and Applications 171 1 pp 311-25

[16] Kireev V S, Guseva A I, Bochkaryov P V, Kuznetsov I A and Filippov S A 2018 Association rules mining for predictive analytics in IoT cloud system Advances in Intelligent Systems and Computing 848 pp 107-12

[17] Kondakov A and Kulik S 2020 Intelligent information system for telemedicine Procedia Computer Science 169 pp 240-3

[18] Yasnitsky L N, Dumler A A and Cherepanov F M 2020 Robot-doctor: what can it be? Advanced Technologies in Robotics and Intelligent Systems (Proceedings of ITR 2019). Mechanisms and Machine Science. Springer, Cham 80 pp 215–22

[19] Nastos P T, Moustris K P, Larissi I K and Paliatsos A G 2013 Rain intensity forecast using artificial neural networks in Athens, Greece Atmospheric Research 119 pp 153-60

[20] Kafedziski V, Pecov S and Tanevski D 2018 Detection and classification of land mines from ground penetrating radar data using Faster R-CNN 26th Telecommunications Forum (TELFOR) (SAVA Center, Belgrade, Serbia)

[21] Nogueira R F, de Alencar Lotufo R and Machado R C 2016 Fingerprint liveness detection using convolutional neural networks IEEE transactions on information forensics and security 11 6 pp 1206-13

[22] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T et al 2018 A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play Science 362 pp 1140-4

[23] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: unified, real-time object detection In Proceedings of the IEEE conference on computer vision and pattern recognition pp 779-88