

Intelligent Dialogue System Based on Deep Learning Technology

Pavlo Kushneryk^[0000-0002-5359-2404], Yuriy Kondratenko^[0000-0001-7736-883X],
Ievgen Sidenko^[0000-0001-6496-2469]

Intelligent Information Systems Department, Petro Mohyla Black Sea National University,
68th Desantnykiv Str., 10, Mykolaiv, 54003, Ukraine,
levelup.kpi@gmail.com, yuriy.kondratenko@chmnu.edu.ua,
ievgen.sidenko@chmnu.edu.ua

Abstract. Recent advances in machine learning has contributed to the rebirth of the chat-bot. Lately we have seen a rise in chat-bot technology being made available on the web and on mobile devices, and recent reports states that 57 % of companies have implemented or are planning to implement a chat-bot in the near future. Chat-bots are therefore a big part of an AI powered future, however recent reviews find chat-bots to be perceived as unintelligent and non-conversational. Such findings have not slowed down the rapid implementation of chat-bots online, and the same mistakes seems to be repeated over and over again. This explains why we need to understand how to develop, deploy and monitoring our own dialog system based on “Deep Learning” technologies. In our case studies we have compared different neural network architectures and develop chit-chat bot which based on encoder-decoder architecture with attention mechanism. In order to achieve this goal we use Python as programming language, TensorFlow as deep learning framework and GoogleNews word embedding. The peculiarities of the “Deep Learning” technology implementation are discussed in detail. Simulation results confirm the efficiency of the proposed approach for speech recognition.

Keywords: neural network, linguistic recognition, machine learning, deep learning, decision making, dialogue system, chat-bot.

1 Introduction

As the world of today is dramatically changing, technology behind the service is also rapidly changing. Technological trends include artificial intelligence and advanced machine learning, virtual and complementary reality, intelligent programs, intelligent things and spoken systems [1-3]. Virtual Personal Assistants, such as Apple's Siri, Google Now, Microsoft Corp., or Amazon Echo, are already making daily tasks easier. In addition to virtual personal assistants, another type of smart application that can improve user experience and make our lives better is Chat Booth [4]. Chat bots can change the way services are rendered. Instead of using websites or installing other

new programs, users could anticipate ordering services through the chat interface [2, 5, 6].

Chat-bot is a computer program or an artificial intelligence which conducts a conversation via auditory or textual methods [7-10]. Such programs are often designed to convincingly simulate how a human would behave as a conversational partner, thereby passing the Turing test. Chat-bots are typically used in dialog systems for various practical purposes including customer service or information acquisition. Some chat-bots use sophisticated natural language processing systems, but many simpler ones scan for keywords within the input, then pull a reply with the most matching keywords, or the most similar wording pattern, from a database:

- A. Dialogue agent: must understand the user, that is, to provide an understanding function. Bots provide text input that is analyzed with a natural language processing tool and is used to create appropriate responses [2].
- B. Rational Agent: must have access to an external database (for example, through the data enclosures) so that it can secure the competence by answering the questions of the users. Keep context-specific information (such as a username, etc.) [2, 4, 5].
- C. Embodied agent: must provide a presence function. This feature is extremely important in the case of custom-made users giving bots the names (ELIZA, ALICE, CHARLIE, etc.) to satisfy this condition. Today, bayonets are focusing on using linguistic tricks to create characters for chat bots to enhance trust in users and create the impression of an incarnate agent [3, 4].

Existing chat rooms, such as Siri, Alexa, Cortana and Google Assistant, face difficulties in understanding user intentions and, therefore, become difficult to work with. In particular, these chat rooms can not track the context and suffer from long-lasting conversations. Another disadvantage of these chat rooms is that they are designed specifically to help the user with some specific problems [7, 8], and therefore, they limit their area. They are not able to make a consistent and engaging conversation between two people on popular topics such as the latest news, politics, and sports.

2 Related Works and Problem Statement

Dialogue system is one of the most difficult areas of artificial intelligence due to the subjectivity associated with the interpretation of the human language. Issues related to the development of the dialogue system include understanding the natural language, presentation of knowledge and dialogue assessment. A complete solution to this problem is likely to have a system of human parity that is difficult to measure. Such systems can be described as AI complete. Based on the latest achievements in Deep Learning and AI, we have received significant performance improvements in AI-full areas such as image recognition and computer vision. These achievements are largely due to the objective nature of evaluating these problems. Dialogue system¹ requires both an understanding of natural language and the generation of responses, where the latter have a potentially unlimited response space, as well as the lack of objective indicators of success, which makes it extremely difficult for modeling [4, 11-13].

Today, there are three main areas for the development of dialog systems: general-purpose dialogue systems (chit-chat), information dialog systems and problem-oriented. Early job-oriented tasks require a large amount of labeled data [6, 7] and are very expensive. Recent work, as a rule, uses deep learning techniques in each component of the dialogue system and demonstrates significant improvements. In researches [12, 13] used the LSTM and the Conditional Network of random fields to fill the slot. Authors in paper [14] extend ANNs of the feed-forward type by using fractional derivative theory. Authors [15] used a multi-armed bandit machine to make decisions. Authors [16] introduced a complex task-oriented dialogue system and a master-shooter to collect data in the restaurant's domain. Authors [17] built a through-the-line target bot based on a memory network. Authors [18] built a target information-based access system based on reinforcement training, trying to highlight related elements with certain attribute values. However, most of these preliminary work focuses on NLP calls instead of commercial success rates, such as conversion rates. They either did not focus on recommendations or did not model or use previous user benefits when recommended by users [19, 20].

The main purpose of this article is to describe the process of developing a combined dialog system for conversation on general topics, and to search the question on the resource "stackoverflow".

3 Architecture of the Dialogue System

The classical scheme of the dialogue system has the form shown in Fig. 1:

1. Natural Language Understanding (NLU) is a natural language recognition module. The main task of this module is identifying entities, identifying the subject of the input sentence and preparing the detected data for further processing [3, 21].
2. Dialogue manager (DL) is the module following the natural language recognition module that aims to coordinate the flow of dialogue and communicate with other subsystems and components [4, 22].
3. Natural Language Generation (NLG) receives a specification of a communicative act from the dialog manager and generates a corresponding text representation. Define the two functions that the response generation module should perform: content planning and language generation, but acknowledge that the first can be attributed to the dialog manager [5].

The natural language recognition module is a very important part of the dialog system, since it depends on which model the data will be accessed in the future, and how the system output will be generated [2-5].

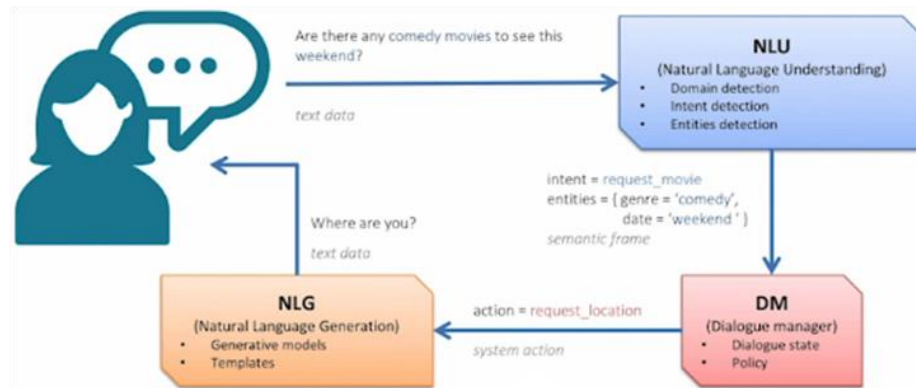


Fig. 1. General scheme of the dialogue system

Depending on the needs of the task the needs of the task and the accuracy required for its solution, this module can be implemented in different ways. Based on the main task, the module can be represented as parser and classifier. The responsibilities of the parser include the identification of significant for the data classifier, which can include named entities, parts of the language, numbers, and so on. In order to solve this problem, the language of regular expressions can be used in systems applications. The advantage of this solution is the speed of interpretation, the main drawback is the complexity of the scaling. Another solution is to use trained models of the type POS (part of speech) to identify parts of languages, and the NER (named entity recognition) system to search for named entities (names, addresses, etc.). The advantage of such models is greater capacity and better quality of prediction, important drawbacks can be attributed to the dependence of the models on the data they are trained and, high computational complexity [23-25].

The Dialog Manager must receive the user input from the natural language learning routine and generate system responses at the level of the concept of the natural language generator. Which answer chooses will depend on the chosen strategy; another aspect of responsibility attributed to the dialogue manager. Strategies are related to preserving the state of conversation and the ability to model the structure of the dialogue outside of one statement. Achieving a flexible dialogue with users should be based on "smart dialog modeling theories and dialogue management". The main tasks of the dialogue manager include: contextual interpretation - the ability to solve ambiguous and reference expressions, management knowledge of the professional field - the ability to drive on a specialized topic and provide access to sources of information. The choice of action is the decision on what to do next. The development or selection of strategies that allow the dialog manager to decide what to say or do based on the current and prior state of affairs [3, 6, 26, 27].

To generate the answer, the seq2seq architecture is chosen for the neural network architecture with the use of the attention mechanism. The use of the mechanism of attention is motivated.

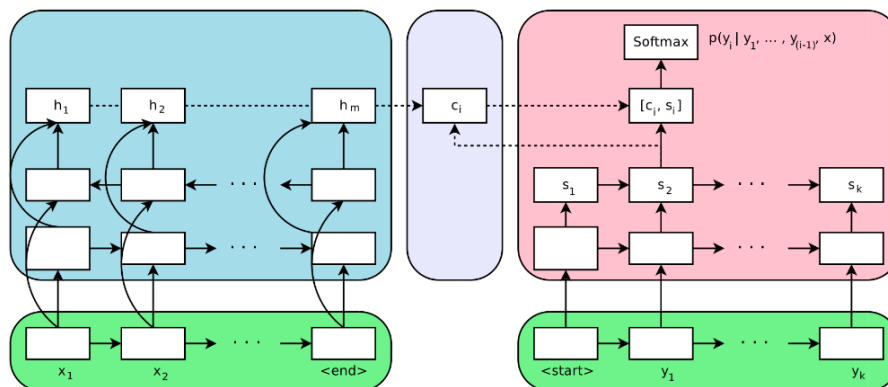


Fig. 2. Seq2seq architecture using the mechanism of attention

Increasing the quality of the system's operation for long sentences Fig. 3. In order to estimate quality of results we have used BLEU score [28]. Main idea of this metrics is: counting matching n-grams in the candidate answer message to n-grams in the reference text, where 1-gram or unigram would be each token and a bigram comparison would be each word pair. The comparison is made regardless of word order.

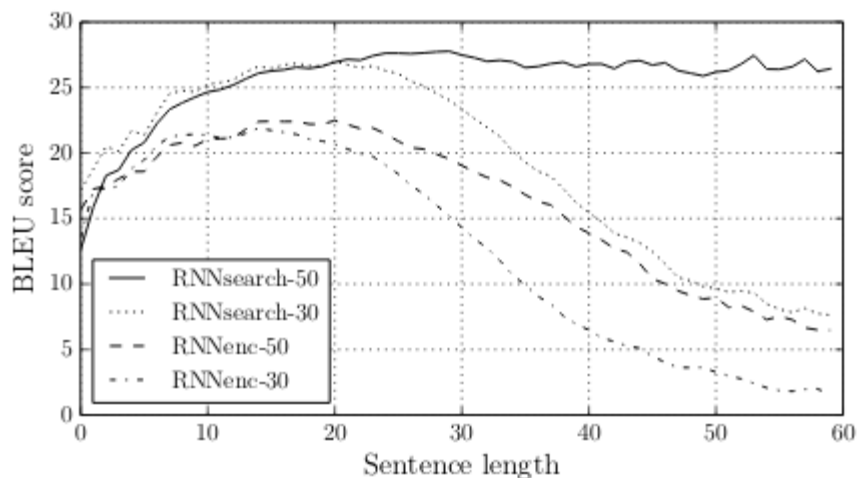


Fig. 3. Dependence of the quality of generation of sentences on length

The main point of the mechanism of attention is the presence of feedback between the encoder and the decoder, which allows for more important words at the input to match the larger weights. The value of weight is calculated as follows:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}, \quad (1)$$

where e_{ij} is an alignment model, which measures how much the inputs at position j and the output around position i are matched; T_x is a length of the input sequence. Values e_{ij} computed as $e_{ij} = a(s_{i-1}, h_j)$, where a is a similarity function.

As the main cell for the network, GRU is selected, which shows better accuracy and smoothness of training than LSTM [28-30].

4 Practical Implementation of the Dialogue System

Graphically, the structure of the dialog system was presented and described in the previous section in Fig. 4. This section is devoted to a more detailed description of the presented diagram. The system described in this publication implements the architecture proposed in the preceding paragraph, here describes in more detail the technologies used in the system, and gives examples of its work [7, 13, 31-33].

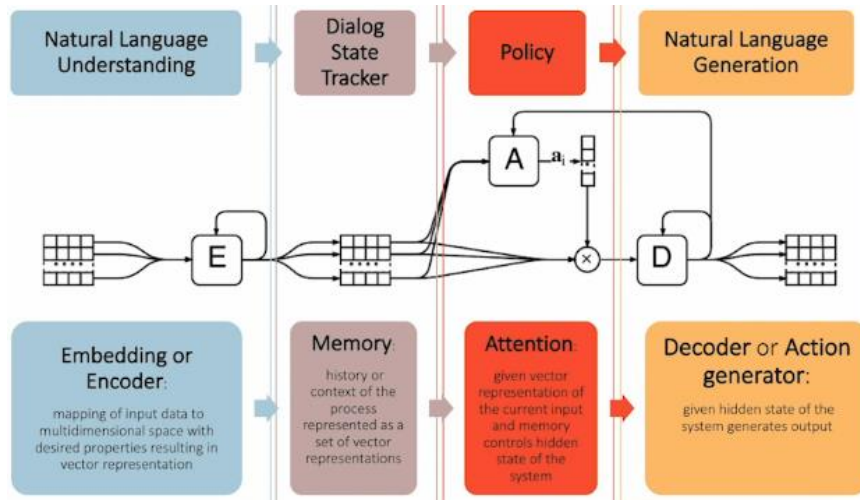


Fig. 4. The structure of the dialogue system in the form of a neural network

The main tasks of the system are to support the dialogue on a free topic, and to find the answer to the question on the “stackoverflow” resource. Proceeding from this, the task of the natural language recognition module is the classification of the incoming sentence as a sentence related to the question, or a sentence related to the dialogue. To implement this node in the dialogue system, the logistic regression of text data is used which involves assigning the text to one of these groups. Then, depending on the results of the natural language recognition module, the sentence can

fall into the search response system for the “stackoverflow” or the dialog manager. The answer search system is implemented as a one-to-one class multi-classifier, which, in the text of the sentence, returns the probable answer to the question. The natural language recognition system, and the classification system, is executed in the Python programming language [7] using the library scikit-learn Dialog Manager (Fig. 4) implemented in the form of seq2seq neural network with a mechanism of attention. The Dialog Manager is implemented in the Python programming language using the Tensorflow framework; Google GloVe is used as a vector representation of words [11, 34, 35].

The dialogue system described in this section is an API that allows integration with different messengers (Telegram, Twiter, etc.) and is used as an independent text processing system. The current dialog system is executed in the form of a Telegram bot, whose task is to support dialogue, and to search for programming information on the “stackoverflow” resource. An example of the system is shown in Fig. 5.

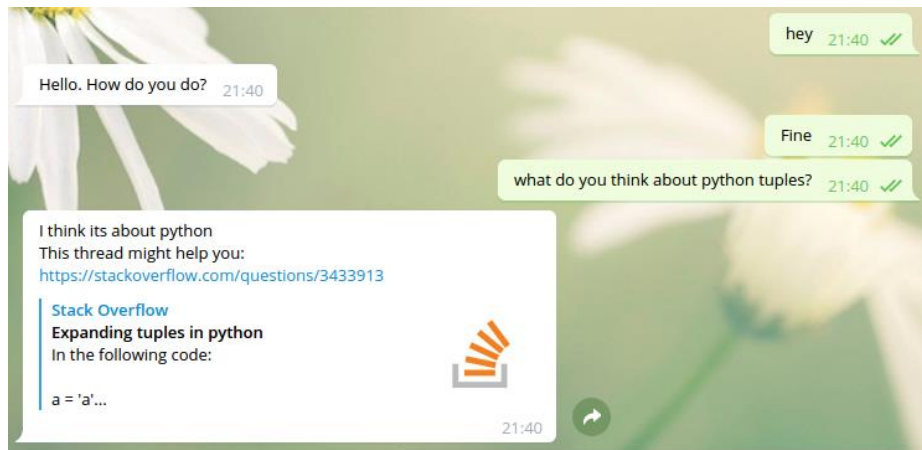


Fig. 5. The result of the operation of the dialogue system

5 Conclusions

The development of dialogue systems is a very important scientific and practical task at the moment. Improving existing and developing new methods for building, training and testing dialogues will help develop in-depth learning on the one hand and improve services on the other.

This paper describes a basic dialogue system using artificial intelligence and machine learning. In our model we use “seq2seq architecture” because it has highly precision on long sentence compared to RNN-Language Model and Feedforward Neural Network Language Model. As basis of more modern architecture seq2seq allows us easily use such solution as “Attention mechanism” in order to improve quality.

In this work authors showed full process of development dialog system using Deep Neural Network from choosing tolls and architecture neural network to examples of

work final system. Also in conclusion parts we motivate pros and cons of chosen neural network architecture.

References

1. Abdel-Hamid, O., Mohamed, A.R., Jiang, H., Deng, L., Penn, G., Yu, D.: Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22(10), 1533-1545 (2014). DOI: 10.1109/TASLP.2014.2339736.
2. Yadav, N., Yadav, A., Kumar, M.: *An Introduction to Neural Network Methods for Differential Equations*. Springer, Netherlands (2015). DOI: 10.1007/978-94-017-9816-7.
3. Zeng, Z., Wang, J.: *Advances in Neural Network Research and Applications*. Springer-Verlag, Berlin, Heidelberg (2010). DOI: 10.1007/978-3-642-12990-2.
4. Anastassiou, G.A.: *Intelligent Systems II: Complete Approximation by Neural Network Operators*. Springer, Switzerland (2016). DOI: 10.1007/978-3-319-20505-2.
5. Da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F.: *Artificial Neural Networks*. Springer, Switzerland (2017). DOI: 10.1007/978-3-319-43162-8.
6. Krawczak, M.: *Multilayer Neural Networks. A Generalized Net Perspective*. Studies in Computational Intelligence 478. Springer, Switzerland (2013). DOI: 10.1007/978-3-319-00248-4.
7. Ketkar, N.: *Deep Learning with Python*. Apress, Berkeley (2017). DOI: 10.1007/978-1-4842-2766-4.
8. Kim, P.: *MATLAB Deep Learning. With Machine Learning, Neural Networks and Artificial Intelligence*. Apress, Berkeley (2017). DOI: 10.1007/978-1-4842-2845-6.
9. Kudermetov, R., Polska O.: Towards a formalization of the fundamental concepts of SOA. In: 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), pp. 492-494. Lviv, Ukraine (2016). DOI: 10.1109/TCSET.2016.7452096.
10. Kondratenko, Y.P., Kozlov, O.V., Gerasin, O.S., Zaporozhets, Y.M.: Synthesis and research of neuro-fuzzy observer of clamping force for mobile robot automatic control system. In: *IEEE First International Conference on Data Stream Mining&Processing (DSMP)*, pp. 90-95. Lviv, Ukraine (2016). DOI: 10.1109/DSMP.2016.7583514
11. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: *Conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543. Doha, Qatar (2014).
12. Li, X., Chen, Y.-N., Li, L., Gao, J., Celikyilmaz, A.: Investigation of Language Understanding Impact for Reinforcement Learning Based Dialogue Systems (2017). arXiv:1703.07055 [cs.CL].
13. Cai, H., Ren, K., Zhang, W., Malialis, K., Wang, J., Yu, Y., Guo, D.: Real-Time Bidding by Reinforcement Learning in Display Advertising. In: *The 10th ACM International Conference on WebSearch and Data Mining (WSDM)*, pp. 661-670. Cambridge, UK (2017). DOI: 10.1145/3018661.3018702.
14. Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P.: From homogeneous network to neural nets with fractional derivative mechanism. In: Rutkowski, L. et al. (eds.) *International Conference on Artificial Intelligence and Soft Computing (ICAISC), Part I, LNCS, vol. 10245*, pp. 52-63. Springer, Cham (2017). DOI: 10.1007/978-3-319-59063-9_5.

15. PATTANAYAK, S.: Pro Deep Learning with TensorFlow. A Mathematical Approach to Advanced Artificial Intelligence in Python. Apress, Berkeley (2017). DOI: 10.1007/978-1-4842-3096-1.
16. Christakopoulou, K., Radlinski, F., Hofmann, K.: Towards Conversational Recommender Systems. In: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 815-824. San Francisco, California, USA (2016). DOI: 10.1145/2939672.2939746
17. Vukovic, D., Dujlovic, I.: Facebook Messenger Bots and Their Application for Business. In: IEEE Transactions on 24th Telecommunications forum (TELFOR), pp. 1-4. Belgrade, Serbia (2016). DOI: 10.1109/TELFOR.2016.7818926.
18. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: International Conference on Learning Representations Workshop (2016). arXiv:1409.0473 [cs.CL].
19. Kondratenko, Y.P., Rudolph, J., Kozlov, O.V., Zaporozhets, Y.M., Gerasin, O.S.: Neuro-fuzzy observers of clamping force for magnetically operated movers of mobile robots. *Technical Electrodynamics* 5, 53-61 (2017). (in Ukrainian).
20. Becker, L., Basu, S., Vanderwende, L.: Mind the gap: Learning to choose gaps for question generation. In: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, pp. 742-751. Montreal, Canada (2012). <http://aclanthology.coli.uni-saarland.de/pdf/N/N12/N12-1092.pdf>.
21. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B.: Adversity-promoting objective function for neural conversation models. In: Proc. of NAACL (2016). arXiv:1510.03055 [cs.CL].
22. Schmidhuber, J.: Deep Learning in Neural Networks: An Overview. *Neural Networks* 61, 85-117 (2015).
23. Kondratenko, Y.P., Kozlov, O.V., Klymenko, L.P., Kondratenko, G.V.: Synthesis and Research of Neuro-Fuzzy Model of Ecoprogenesis Multi-circuit Circulatory System. In: Jamshidi, M., Kreinovich, V., Kazprzyk, J. (eds.) *Advance Trends in Soft Computing, Series: Studies in Fuzziness and Soft Computing*, vol. 312, pp. 1-14. Springer, Cham (2014). DOI: 10.1007/978-3-319-03674-8_1.
24. Bengio, Y.: Learning deep architectures for AI. *Foundations and Trends in Machine Learning* 2(1), 1-127 (2009). DOI: 10.1561/2200000006.
25. Britz, D.: Recurrent neural networks tutorial, Part 1 – Introduction to RNNs (2017). <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>.
26. Solesvik, M., Kondratenko, Y.: Architecture for Collaborative Digital Simulation for the Polar Regions. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) *Green IT Engineering: Social, Business and Industrial Applications. Studies in Systems, Decision and Control*, vol. 171, pp. 517-531. Springer, Cham (2019). DOI: 10.1007/978-3-030-00253-4_22
27. Kondratenko, Y., Gordienko, E.: Implementation of the neural networks for adaptive control system on FPGA. In: *Annals of DAAAM for 2012 & Proceeding of the 23th Int. DAAAM Symp. Intelligent Manufacturing and Automation*, vol. 23(1), pp. 0389-0392 (2012).
28. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machinetranslation. *InACL* (2002).
29. Bengio, Y., LeCun, Y., Hinton, G.: Deep Learning. *Nature* 521, 436-44 (2015).

30. Kondratenko, Y., Gordienko, E.: Neural Networks for Adaptive Control System of Caterpillar Turn. In: Annals of DAAAM for 2011 & Proceeding of the 22th Int. DAAAM Symp. Intelligent Manufacturing and Automation, pp. 0305-0306 (2011).
31. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of online learning and an application to boosting. *J. Comput. Syst. Sci.* 55(1), pp. 119-139 (1997).
32. Xiong, Y., Zuo, R.: Recognition of geochemical anomalies using a deep autoencoder network. *Comput. Geosci.-UK* 86, 75-82 (2016).
33. Liu, P., Han, S., Meng, Z., Tong, Y.: Facial Expression Recognition via a Boosted Deep Belief Network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Columbus (CVPR), pp. 1805-1812 (2014).
34. Bahdanau, D., Cho, K., Bengio, Y. Neural machine translation by jointly learning to align and translate (2014). arXiv:1409.0473.
35. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proc. Neural Information and Processing Systems, pp. 1097-1105 (2012).