Development of the architecture of a transformer-based neural network model to automate delivering judgments in bankruptcy cases

Petr Pylov\(^1\)*, Roman Maitak\(^1\), Anna Dyagileva\(^1\), and Andrey Protodyakonov\(^1\)

\(^1\)T.F. Gorbachev Kuzbass State Technical University, 650000, Kemerovo, Russia

Abstract. Delivering judgments is one of the brightest examples of solving a creative problem, which implies not only the analysis of data presented in natural language, but also the verification of the compliance of the input information with legal norms and rules. Automation of this process requires the creation of such a language model of machine learning that would allow processing natural language and delivering judgments based on the legal framework, thereby completely replacing the position of a judge. Serious functional requirements are imposed on such an intelligent system, which describe the system of constraints for the architecture of a machine learning model in a formalized mathematical language. This article is devoted to defining the rules for building an applied artificial intelligence model that would automate the process of delivering judgments in bankruptcy cases.

1 Introduction

After the active integration of the credit system into the global economy, the opportunities for buyers and sellers expanded: the former could afford to purchase goods or services immediately, without postponing the purchase for a long time. The latter, by attracting additional financial resources, have discovered the way for themselves to increase the turnover of goods and services that they offer their customers.

However, it is worth noting that with the growth in the number of opportunities, the responsibility borne by the subject of the economy will naturally increase. Those market participants who failed to reasonably balance their opportunities with their responsibilities often suffered an economic collapse, which is often referred to as bankruptcy. If it is impossible to pay the bills, the company itself (or an individual) or its creditors go to court to initiate bankruptcy proceedings - the legal recognition of a state of insolvency. In accordance with local law, the court accepts a petition in bankruptcy and begins the hearing of the case. The logical conclusion of this process is the issuance of a court decision, which may be different in each specific situation, but must necessarily be based on the current legal framework.

* Corresponding author: pylovpa@kuzstu.ru

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
Automating the process of delivering judgment in bankruptcy cases is an urgent task, as it allows reducing the burden on a judge or completely replace him in the process of delivering judgments. Undoubted advantages of automation are:

- Reducing the burden on a judge. This is especially important in cases where the state body receives a large number of petitions from various subjects of the market economy;
- Reduced chance of human error. It is common for any person to make mistakes, therefore, despite the possibility of appealing against an erroneous court decision, the option of fully eliminating this kind of error will always be a priority;
- Speeding up a judgment. Having a computer center in the courthouse would allow the intelligent system to process each petition hundreds of times faster than if the decision was delivered by a person. In this case, even the criterion of a person's skills is insignificant (it can be a novice judge or a judge with dozens of years of experience), since the judge spends time reading, analyzing the document, making a judgment on the basis of legislation and, finally, formalizing it in the form of a text document of a court decision. From the point of view of the functioning of the machine learning model for natural language processing, all these processes are performed in parallel and almost simultaneously [1], which leads to a huge difference in the time required to deliver a judgment;
- Significant reduction in government spending. Judges are paid high salaries, and they are also entitled to a wide range of benefits and incentives, which is a significant item of expenditure for the budget. The cost of an intelligent system is reduced to the purchase of a computer center on the basis of which it will function;
- The process of constant training and knowledge building is also automated in the intellectual model. Undoubtedly, the professionalism of a judge is determined by his skills and knowledge of the laws that he applies in practice. However, today's society is constantly changing, and judicial precedents are constantly emerging, which every judge must study. In addition, the legislative framework is periodically adjusted, and amendments and changes are made to the laws, which also need to be studied by judges. It is obvious that an intelligent system adapts to innovations much faster and with less inertia, since it only requires uploading amendments or judicial precedents to the general database, without requiring additional resources (hardware and time) to update its “knowledge base” [2].

Having determined the advantages of the automated system, we note that it can also act as an expert system to assist judges in delivering judgments.

It is also worth noting that the intelligent system also has drawbacks. One of the main disadvantages is the accuracy and completeness of the judgment, which directly depends on the architecture of the machine learning model. Another disadvantage is the potential for an attack on such a system in order to deliver a given judgment in favor of the interested party of the case. The second disadvantage can be eliminated if the system operates in a local computer center, which will not have access to the Internet and will be protected from physical attack (for example, it will be guarded around the clock). The first drawback is eliminated by the development of a special architecture of the language model, which will be considered in this article.

2 Results and discussion

The solution to the adjudication problem directly depends on how well the deep learning model will process natural language. This is one of the creative tasks that can be solved in machine learning using the transformer model [3].

To solve the problem of automating judgment delivery based on the architecture of transformers, the authors of this article consider the following heuristics (Figure 1).
The logic of the analysis of input data (petitions) is shown in Figure 1.

The "W" set is the input data set, which is the sentences of the petition written in natural language and arranged in the order in which they appear in the document. Further, all source words are translated into embeddings, as a result of which transition number 1 is performed, and the set "W" is mapped to the set "E".

The second transition (from the set "E" to the set "C") is provided due to the presence of a transformer (attention models placed sequentially one after another for information processing). This transition provides one of the most important functions in the task of processing petitions: it will encode the vectors of the set "E" into the vectors of the set "C", which will provide information about the context of each word in the sentence [4]. This transformer acts as an encoder and will allow translating the vectors of the original words into some hidden vectors that correctly store information about the context of the words. The architecture of the transformer encoder (shown in Figure 2) is based on multi-head attention models, end-to-end connections, and normalizations [5].

The set "C" is the result of encoding the input sequence. However, our goal was to deliver a judgment, so in order to train the model, the algorithm will first have to consider the outcome of the case - the judgment that the judge delivered. Thus, through the third transition, it is necessary to perform decoding in order to translate the input sequence (petition) into the output sequence of sentences (judgment), which will consist of embeddings of the words of the output language. In other words, the third transition allows transforming the original sequence of the set "C", which has the length "n" into the sequence of the set "O", which will have the length "m".

Sequence lengths are given different letters because they will almost always be different: the judge is not required to deliver a judgment equal in number of words to the petition. The architecture of the transformer decoder is arranged almost similarly to the model of the transformer encoder, the only difference is that the masked representation of the context of all the words of the encoder is fed into the middle of the block of the transformer decoder (Figure 3). Thus, only the last encoder block provides information to all transformer decoder blocks [5].
Fig. 2. The logical architecture of the transformer encoder

Fig. 3. The logical architecture of the transformer decoder (on the right) linked to those of the transformer encoder (on the left)
Finally, the fourth transition will convert the embeddings into output natural language sentences that can be read by humans and understood by court paper flow specialists. In a formalized mathematical language, this process can be described as the transformation of the vectors of the set "O" into words (replacing vectorized representations) of the set "Y". Probabilistic language models are widely used for this transition [5].

3 Conclusion

The architecture of the machine learning model described in this article allows using all the advantages of transformer models in order to deliver the most accurate and complete judgment.

The idea of concatenating an encoder transformer and a decoder transformer in one resulting architecture of a machine learning model makes it possible to achieve the accuracy and completeness of the solution comparable to transfer learning models[6]. However, the advantage of the architecture considered in the article over transfer models is that there is no need to use third-party data, which are required to increase the accuracy and completeness of the approximation. This advantage allows data scientists to focus on solving a specific problem (delivering judgments) without involving third-party data for the process of training a machine learning model. In the future, this will make it possible to abstract from additional data research (for example, natural language processing as such) and focus only on replenishing the database of an intelligent system [7] with data that are directly related to solving an applied problem: adding new judicial precedents and updating the legislative framework.

References

5. J. Brownlee, Machine Learning Al (2016)