

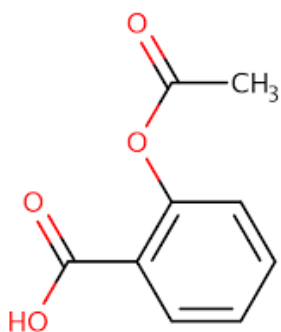
Deep learning from structured data

During the accelerated development of telecommunications, networking, finance and industry, a large amount of data is generated continuously. We can process and analyze them using machine learning methods. Deep neural networks are nowadays the most important class of learning models, which are able to solve complex problems. However, in its basic form, neural networks can only process data represented as vectors of the same size.

In practice, we very often encounter structured data which cannot be directly represented in this form:

- In natural language processing, documents consist of sequences with different lengths.
- In cheminformatics, chemical compounds are represented by graphs.
- In social sciences, data are represented as complex networks.

Since learning from structured data cannot be performed automatically with use of typical deep models, additional tools have to be designed. Every new data type causes nontrivial problems of its input into a neural network, which is problematic for practitioners. In our opinion, current methods are insufficient to preserve the whole information contained in complex description of structured objects.



(a) Graph of chemical compound.



(b) Network representing complex relationships

The aim of this project is to construct a theoretically justified *general framework* to enable in a uniform way the neural network to process practically every possible data-type. The research objective is consistent with current trends in present machine learning and, in consequence, can have an impact on the scientific discipline. In particular, the output of the project will also have direct consequences in natural language processing and computer vision. These are two most important areas for application of deep learning.

In this project, we will construct a generic approach, which can be applied to different types of structured data. We will give theoretical justifications for the proposed methods, which will allow to better understand our methods. Moreover, we will verify its usefulness on various practical problems, such as documents classification, image reconstruction, chemical activity prediction.