Application of optimization methods in solving problems of chemoinformatics

Andrianova A A, Mukhin N V

Kazan (Volga Region) Federal University

Abstract. The aspects of solving the chemical reaction mapping problem using the Support Vector Machine method and the problem of quadratic convex minimization are considered.

In this message one of problems of chemoinformatics is considered - obtaining chemical reaction mapping, i.e. search for the correspondence of atoms of the left and right parts of the chemical reaction ([1]).

For a mathematical model it is assumed that the chemical reaction is a sequence of operations on the atoms of the left part of the chemical reaction, as a result of which the right part is obtained. Some basic operations are removing the bond from an atom, inserting a bond from an atom, completely detaching or attaching an atom to another are introduced. As a vector describing the reaction, we will use a vector with components-quantities of significant chemical transformation operations. So, we consider the problem of classifying the mappings of a chemical reaction into two classes — correct and incorrect.

To solve this binary classification problem the Support Vector Machine method was applied, which reduces to the optimization problem with following dual problem:

$$\min\left\{\sum_{i=1}^{l}\lambda_i - 0.5\sum_{s=1}^{l}\sum_{t=1}^{l}\lambda_s y_s \lambda_t y_t \left\langle x_s, x_t\right\rangle, 0 \le \lambda_i \le C, i = 1\ldots l\right\},$$

where $l > 0$ — number of examples in the training sample, $(x_i, y_i)(i = 1\ldots l)$ is an example with a class label $y_i (y_i \in \{-1, 1\})$ — correct or incorrect mapping, λ_i $(i = 1\ldots l)$ — a set of dual variables along which one can obtain the vector of the coefficients of the separating hyperplane $(\sum_{i=1}^{l}\lambda_i x_i y_i)$.

To solve this dual problem, an algorithm was applied with the decomposition to the construction of coordinate-wise descent algorithms, proposed in [2].

References

1. William Lingran Chen, David Z. Chen and Keith T. Taylor: Automatic reaction mapping and reaction center detection. WIREs Comput Mol Sci 3, 560-593 (2013).
2. Konnov I.V.: Sequential threshold control in descent splitting methods for decomposable optimization problems. Optimization Methods and Software. V.30, No 6, 1238–1254 (2015).