

ПРИМЕНЕНИЕ НОВОГО КОЛЛЕКТИВНОГО АЛГОРИТМА СТАЙНОГО ТИПА ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ МЕТОДОМ ОПОРНЫХ ВЕКТОРОВ

Ахмедова Ш.А.

Научный руководитель д-р техн. наук Семенкин Е. С.

Сибирский Государственный Аэрокосмический Университет им. академика М.Ф. Решетнева

Метод опорных векторов (Support Vector Machine, SVM) – алгоритм вида «обучение с учителем», использующийся для задач бинарной классификации и регрессионного анализа [1]. Данный метод может быть применен для решения различного рода задач, например, задач текстовой категоризации, медицинской диагностики, банковского скоринга и т.д.

Сама по себе задача классификации методом опорных векторов в простейшем варианте сводится к решению задачи условной оптимизации, которая описывается следующим образом. Пусть X – пространство объектов, $Y = \{-1; 1\}$ – классы. Дана обучающая выборка: $(x_1, y_1), \dots, (x_m, y_m)$. Требуется построить разделяющую объекты из разных классов гиперплоскость, удовлетворяющую условиям:

$$\begin{aligned} \arg \min_{w,b} \|w\|^2, \\ y_i * \text{sign}(\langle w, x \rangle + b) \geq 1, i = 1, \dots, m, \end{aligned}$$

где $\langle \dots \rangle$ – скалярное произведение, w – нормальный вектор к разделяющей гиперплоскости, b – вспомогательный параметр.

Однако точность классификации может меняться в зависимости от выбора алгоритма оптимизации. В настоящее время существует множество алгоритмов, с помощью которых можно находить решения широкого класса задач оптимизации с любой заранее заданной точностью (алгоритмы стайного интеллекта, генетические алгоритмы и т.д.). Эффективность применения этих алгоритмов слабо зависит от вида оптимизируемого функционала, к тому же их несложно реализовать.

Исследование эффективности различных алгоритмов стайного интеллекта (Particle Swarm Optimization [2], WolfPackSearch [3], FireflyAlgorithm [4], CuckooSearchAlgorithm [5] и Bat Algorithm [6]), имитирующих групповое поведение некоторых популяций в природе, показало целесообразность их применения, хотя и у этих оптимизационных методов есть свои недостатки. Во-первых, эффективность их работы определяется настройкой различных параметров, в частности – размера популяции, что тоже является весьма сложной оптимизационной задачей. И, во-вторых, исследования показали, что нельзя заранее определить, какой именно алгоритм стоит применять для решения той или иной задачи оптимизации: для различных задач (и даже для различных размерностей одной и той же задачи) лучшие результаты достигаются различными методами.

Поэтому был разработан новый самонастраивающийся алгоритм оптимизации на основе перечисленных выше эвристик, названный Co-Operation of Biology Related Algorithms (COBRA) [7]. Главная идея разработанного подхода заключается в генерировании пяти популяций (одной для каждой эвристики), которые в дальнейшем коллективно решают задачу оптимизации на основе конкуренции и кооперации. Исследование эффективности разработанного метода продемонстрировало его работоспособность, а также то, что он показывает лучшие результаты, чем все перечисленные алгоритмы-компоненты по отдельности [7].

Далее предложенная эвристика была модифицирована для решения задач условной оптимизации. Модифицированный алгоритм был также тщательно проверен на репрезентативном множестве тестовых задач условной оптимизации [8]. В данной работе эта модификация используется для настройки параметров метода опорных векторов при решении задач классификации.

Методом опорных векторов, настраиваемым алгоритмом COBRA для решения задач условной оптимизации, были решены две задачи банковского скоринга: скоринг по базе данных из Германии и скоринг по базе данных из Австралии [9]. Данные задачи обычно используются разными специалистами для оценки разрабатываемых методов классификации и сравнения их с известными подходами. В частности, ранее эти задачи были решены с помощью достаточно простых нейроклассификаторов, настраиваемых алгоритмом COBRA для безусловной оптимизации [10]. Структура нейросетей была установлена в виде полносвязного перцептрона с тремя и пятью нейронами на одном скрытом слое и биполярной функцией активации, COBRA была использована для настройки весовых коэффициентов нейросетей. Результаты сравнения эффективности (в данном случае «эффективность» означает долю правильно классифицированных данных) работы алгоритма классификации, основанного на методе опорных векторов, настраиваемом алгоритмом COBRA, представлены ниже. Выделены результаты, полученные в данной работе методом опорных векторов и нейросетями в работе [10].

Классификатор	Скоринг в Австралии	Скоринг в Германии
2SGP	0.9027	0.8015
SVM+COBRA	0.9022	0.7960
C4.5	0.8986	0.7773
Fuzzy	0.8910	0.7940
ANN+COBRA(5)	0.8907	0.7829
ANN+COBRA(3)	0.8898	0.7809
GP	0.8889	0.7834
CART	0.8744	0.7565
LR	0.8696	0.7837
CCEL	0.8660	0.7460
RSM	0.8520	0.6770
Bagging	0.8420	0.6840
Bayesian	0.8420	0.6790
Boosting	0.7600	0.7000
k-NN	0.7150	0.7151

Таким образом, установлено, что предложенный способ классификации методом опорных векторов, настраиваемым алгоритмом COBRA, показывает лучшие результаты, чем большинство альтернативных методов классификации. Это означает, что новый алгоритм COBRA для условной оптимизации успешно прошел апробацию на сложной практической задаче.

Литература

1. Воронцов К.В. Лекции по методу опорных векторов (режим доступа: www.machinelearning.ru).
2. Kennedy J., Eberhart R. Particle Swarm Optimization. // Proceedings of IEEE International Conference on Neural Networks. – 1995. – IV. – P. 1942-1948.

3. Yang Ch. Algorithm of Marriage in Honey Bees Optimization Based on the Wolf Pack Search // Proceedings of International Conference on Intelligent Pervasive Computing. – 2007. – P. 462-467.
4. Yang X.S. Firefly algorithms for multimodal optimization // Lecture Notes in Computer Science. – 2009. – P. 169-178.
5. Yang X.S., Deb S. Cuckoo search via Levy flights // Proc. of World Congress on Nature & Biologically Inspired Computing. – 2009. – P. 210-214.
6. Yang X.S. A new metaheuristic bat-inspired algorithm // Nature Inspired Cooperative Strategies for Optimization. – 2010. – P. 65-74.
7. Akhmedova Sh., Semenkin E. Co-Operation of Biology Related Algorithms // Proc. of IEEE Congress on Evolutionary Computation. – 2013.
8. Akhmedova Sh., Shabalov A. Development and Investigation of Biologically Inspired Algorithms Cooperation Metaheuristic // Proceedings of Genetic and Evolutionary Computation Conference. – 2013.
9. A. Frank, A. Asuncion. UCI Machine Learning Repository [режимдоступа: www.archive.ics.uci.edu/ml].
10. Ахмедова Ш.А., Семенкин Е.С. Коллективный метод оптимизации на основе стайных бионических алгоритмов // Системный анализ и информационные технологии: Труды V Международной конференции. – 2013. – Т. 2. – С. 322-331.