

0.1. Полонская Я.С. Определение эффективности методов интеллектуального анализа данных в исследовании природных сигналов

Исследование ионосферы является актуальной задачей для многих областей человеческой деятельности. Ее исследование позволит повысить надежность работы навигационных систем дальней радиосвязи, обеспечения полётов космических аппаратов и т. д. Не менее актуальной темой так же является и прогнозирование сейсмических толчков так как это важнейшее мероприятие в системе контроля сейсмической обстановки, позволяющее своевременно принять меры к защите населения и территорий, резко снизить затраты на ликвидацию последствий землетрясений.

Для проведения исследований были использованы две базы данных. Первая из них соответствует задаче классификации высокочастотных радарных сигналов, возвращаемых из ионосферы. База данных содержит 351 экземпляр и 34 атрибута [1]. Вторая задача состоит в прогнозировании сейсмических толчков. База данных содержит 2584 экземпляра и 19 атрибутов [2]. Задачи решались при помощи программной системы анализа данных RapidMinerStudio [3].

В качестве методов анализа данных для обоих наборов данных были выбраны следующие алгоритмы: искусственная нейронная сеть (ANN), индукция правил (RI), метод k ближайших соседей (k-NN), дерево решений (DT), наивный Байесовский классификатор (NB), случайный лес (RF), логистическая регрессия (LR), метод опорных векторов (SVM).

Для первого набора данных лучшие результаты показали DT- 92.88% и ANN- 91.84%. После оптимизации, которая улучшила показатели всех методов, лучший результат было получен нейронной сетью (ANN98.21%). Применение ансамблей привело к ухудшению результатов.

Для второго набора данных лучший результат показала ANN- 92.6% и RF- 92.8%. Оптимизация параметров также дала улучшение для всех методов, а лучший результат продемонстрировала нейронная сеть ANN(96.94%). Ансамбли обоих типов также не привели к улучшению результатов.

Таким образом, мы наблюдаем, что в обоих случаях из всех использованных подходов лучший результат показала оптимизированная по скорости обучения нейронная сеть. Можно сделать вывод, что задачи исследования природных процессов могут быть решены с помощью программной системы RapidMiner, что может облегчить работу многих организаций, выполняющих данные исследования.

В дальнейшем предполагается перейти от использования стандартных пакетов к применению более эффективных методов, основанных на специальных

эволюционных алгоритмах [4, 5].

Список литературы

- [1] IonosphereDataSet [Электронный ресурс]. URL: <http://archive.ics.uci.edu/ml/datasets/Ionosphere> (датаобращения: 5.09.2015).
- [2] Seismic-bumpsDataSet [Электронный ресурс]. URL: <http://archive.ics.uci.edu/ml/datasets/seismic-bumps> (датаобращения: 6.09.2015).
- [3] RapidMiner [Электронный ресурс]. URL: <https://rapidminer.com/> (дата обращения: 6.09.2015).
- [4] БЕЖИТСКИЙ С.С., СЕМЕНКИН Е. С., СЕМЕНКИНА О. Э. Гибридный эволюционный алгоритм для задач выбора эффективных вариантов систем управления // Автоматизация. Современные технологии. — 2005. — № 11. С. 24.
- [5] СЕМЕНКИН Е. С., СЕМЕНКИНА М. Е. Применение генетического алгоритма с модифицированным оператором равномерной рекомбинации при автоматизированном формировании интеллектуальных информационных технологий // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. — 2007. — № 3 (16). С. 27-33.