

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ ОТДЕЛЬНЫХ ВИДОВ ЗОНТИЧНЫХ ФЛОРЫ КРЫМА

Гаврилова Т.М., Орлов М.А.¹

Московский государственный университет имени М.В. Ломоносова, биологический факультет, кафедра геоботаники, Россия, 119234, Москва, Ленинские горы, д. 1, стр. 12, (495) 939-50-21, gavrilova.t.m@list.ru

¹ Институт биофизики клетки Российской академии наук, Россия, 142290, г. Пущино Московской области, ул. Институтская, д. 3, (4967) 73-91-65, orlovmikhailanat@gmail.com

Моделирование распространения видов позволяет установить связь между местонахождениями особей вида и факторами природной среды. Обработка подобных массивов данных определяет необходимость применения методов машинного обучения. В данной работе ряд таких методик были использованы для моделирования распространения отдельных видов Зонтичных флоры Крыма.

Данные о распространении растений получены на основе этикеток гербарных образцов из трех гербариев: Московского университета (MW), Главного ботанического сада имени Н.В. Цицина РАН (МНА), Ботанического института имени В.Л. Комарова РАН (LE). В качестве факторов среды для моделирования использованы данные WorldClim версии 2.0 [1]. Значения для 19 климатических переменных были извлечены для точек в узлах решетки разрешением 0.125°. Статистический анализ, машинное обучение и визуализация результатов выполнены в свободной программной среде R.

Для биоклиматических переменных проведен анализ главных компонент (РСА), позволивший оценить корреляцию между переменными и их относительный вклад в дисперсию данных. Кластерный анализ (методом Уорда и k -средних) позволил выделить 5 устойчивых кластеров. Результаты кластеризации были сопоставлены с существующими схемами физико-географического районирования Крыма.

С использованием алгоритмов машинного обучения (наивного байесовского классификатора и классификатора “Случайный лес”) были созданы модели распространения отдельных видов Зонтичных флоры Крыма. Для тренировки моделей использовали биоклиматические данные в точках сбора образцов и в случайно выбранных точках решетки. Для отбора оптимальных параметров машинного обучения в разных реализациях использованы различные соотношения этих групп данных.

Литература

1. *Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, A. Jarvis* Very high resolution interpolated climate surfaces for global land areas *International // Journal of Climatology*, 25, 2005. Pp. 1965-1978.