

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ АЛГОРИТМА НЕЙРОННОЙ СЕТИ ПРИ ПРОВЕДЕНИИ ФИЛОГЕНЕТИЧЕСКОГО АНАЛИЗА БИОЦЕНОЗОВ И ОПРЕДЕЛЕНИИ УСТОЙЧИВОСТИ АГРОЛАНДШАФТОВ

А.В. Линкина, А.И. Ярнык

Воронежский институт высоких технологий, г. Воронеж, Россия

Изучение проблем бета-разнообразия (beta-diversity) по Р. Уиттекеру, определяющего степень дифференцированности видов по градиентам местообитания, представляет существенный научный и практический интерес при выявлении морфобиологического состава агроландшафтов с целью обеспечения фенотипического разнообразия для повышения устойчивости сельскохозяйственных угодий. Отдельного внимания заслуживает проблема изучения микробиома почвы в контексте применения диагностических методов оценки уровня генетического разнообразия для формирования адаптационного потенциала видов с учетом минимизации потерь биологического разнообразия в результате антропогенной деятельности человека, а также изменения климатических и других абиотических факторов.

Применение современных инструментов, основанных на машинном обучении, позволяет эффективно решать задачи, связанные с выявлением градиентов на уровне популяции. Авторами были изучены возможности применения алгоритма нейронной сети LMdist (Susan L. Hoops et. All), решающего проблемы коррекции перенасыщенных мер парного расстояния и несходства при проведении филогенетического анализа биоценозов в зависимости от условий среды. Установлено, что при визуализации многомерных наборов данных взаимосвязи микробиома и фитоценоза с условиями среды, могут возникать аномалии, наиболее распространёнными из которых являются т.н. «арки» и «подковы». Кроме того, рядом исследователей (Новаковский А.Б., Дубровский Ю.А.) среди недостатков прямой ординации отмечается отсутствие возможности учета совместного действия факторов среды (кроме канонического анализа соответствий, СА). В отличие от последней непрямая ординация визуально показывает вариабельность данных, существующие в них структуры и тренды. Существует несколько основных видов непрямой ординации, среди которых полярная ординация, анализ главных компонент, анализ соответствий, смещенный анализ соответствий, неметрическое многомерное шкалирование, реципрокное взвешивание. Нами были рассмотрены возможности применения механизма кластеризации при работе алгоритма реципрокного взвешивания и интеграция отмеченного ранее алгоритма LMdist для устранения аномалий.

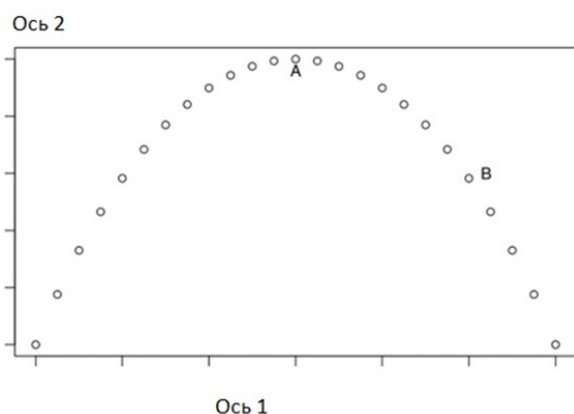


Рисунок 1. – Аномалии градиента в виде «подковы»

Проиллюстрируем применение алгоритма на следующем примере. В общей теории ординации визуализация основана на графическом представлении размещения эколого-ценотических групп по осям в качестве которых может выступать большое разнообразие средовых факторов (температура, влажность, увлажненность, давление, загрязненность химическими веществами и радионуклидами, засоленность, геохимический состав почв, темпоральные характеристики и т. п.). Одна из осей на полученном графике будет представлять влияние фактора, находящегося в максимуме. Однако, вторая и последующие оси могут быть зависимыми от

первой и отражать квадратичные связи между ними. Тогда графически она будет иметь представление в виде измененного графика, получившего название «подковы» (рис. 1).

Аналогично аномалия в виде «арки» выглядит как отображение образцов градиента в виде изогнутой линии (рис. 2.). Указанные аномалии искажают результаты анализа данных бета-разнообразия, поскольку не отражают истинных значений градиента при фенологическом анализе.

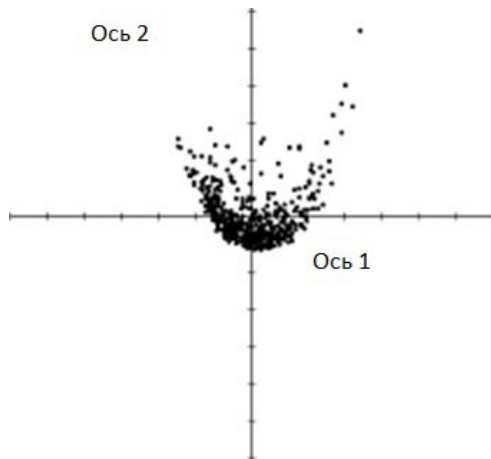


Рисунок 2. – Аномалии градиента в виде «арки»

Таким образом, парные результаты искажают общую картину и свидетельствуют о ложных зависимостях генетического разнообразия фитоценозов, выявляемых при ранжировании фенотипов в зависимости от условий среды.

Поскольку показатели бета-разнообразия могут иметь ограниченный динамический диапазон при отображении больших градиентов окружающей среды с высокой сменой видов, то применение алгоритма позволяет проецировать попарные расстояния между осями. Это позволяет использовать значения, выходящие за пределы диапазона исходного измерения. Не все наборы данных будут иметь перенасыщенные попарные расстояния, поэтому LMdist корректирует только те значения, которые могут быть занижены при наличии выбранного градиента. Скорректированные

расстояния служат исходными данными для определения координат и статистического тестирования. Кроме того, алгоритм позволяет выявлять вариации, ортогональные градиенту. Фрагмент листинга алгоритма на языке Python с использованием стандартных библиотек предварительной подготовки и анализа данных (sklearn) и визуализации (matplotlib) представлен на рисунке 3.

```
2 from sklearn import manifold, datasets
3 from matplotlib import cm
4 from matplotlib.colors import Normalize, to_hex
5 import numpy as np
6
7 # create dataset
8 sr_pts, sr_color = datasets.make_swiss_roll(300, random_state=0)
9
10 # visualize dataset (3D)
11 fig = plt.figure(figsize=(8, 6))
12 ax = fig.add_subplot(111, projection='3d')
13 fig.add_axes(ax)
14 ax.scatter(sr_pts[:,0], sr_pts[:,1], sr_pts[:,2], c=sr_color, s=50, alpha=0.9, cmap="plasma")
15 ax.set_title("Swiss Roll 3D", size=16, fontweight="bold", fontname="Arial", x=0.25, y=0.9)
16 ax.view_init(azim=-75, elev=10)
17 _ = ax.text2D(0.15, 0.71, s="n samples = 300", transform=ax.transAxes, fontname="Arial", size=12)
18 plt.savefig('swiss_roll_3d.png', dpi=1200)
19 plt.savefig('swiss_roll_3d_lowres.png', dpi=350)
20 plt.show()
```

Рисунок 3. – Фрагмент листинга алгоритма LMdist

Таким образом, предложенные методы позволяют отразить влияние экологических факторов на фенологический состав фитоценоза, уточнить построенные градиенты и устранить влияние аномалий за счет применения алгоритма, упорядочивающего внутреннюю структуру используемых наборов данных. Использование инструментов машинного обучения позволяет автоматизировать вычисления и обрабатывать значительные объемы

данных, что является особенно важным на данном уровне развития биоинформатики и фитоценологии.

Работа выполнена при грантовой поддержке Федерального агентства по делам молодёжи (Росмолодёжь) Соглашение № 091–10–2023–069 от 23.05.2023 г. проект «Наука рядом».

Литература

1. Пузаченко ЮГ Математические методы в экологических и географических исследованиях. М. 2004. – 416 с.
2. Новаковский АБ Методы ординации в современной геоботанике. Вестник института биологии Коми научного центра Уральского отделения РАН. 2008; 10 (132):2–8.
3. Новаковский АБ Существующие программные решения для обработки геоботанических данных. Молодежь и наука на Севере. Мат. докл. II Всероссийской (XVII) молодежной научной конференции (с элементами научной школы), 2013:101–104.
4. Susan L Hoops, Dan Knights, LMdist: Local Manifold distance accurately measures beta diversity in ecological gradients. Bioinformatics.2023; № 39(12):72