## Аннотация

Современный мир все больше придает значение экологии и изучению сложных экосистем. Традиционные аналитические методы не всегда способны полностью объяснить и предсказать динамику биологических систем из-за их сложности и разнообразия. В таких случаях компьютерные симуляции являются мощным инструментом для исследования и моделирования экологических процессов. Цель данного проекта - разработка и проведение компьютерных симуляций, основанных на моделировании стационарных сообществ. Проект включает разработку специализированного пакета для симуляций, проведение большого числа симуляций и обработку данных. Модель Дикмана-Лоу выбрана для проведения симуляций, так как она учитывает стохастические процессы и случайные факторы в динамике популяций. Она также позволяет учесть взаимодействия между индивидами и пространственную структуру сообщества.

Ключевые слова: стохастические симуляции, IBM модели (Individual Based Models), компьютерные симуляции, математическая биология, математическое моделирование, модель Дикманна-Лоу.

[**Аннотация 1**](#_qmyjtclu411x)

[**Введение 2**](#_ep092gftmqy)

[Основные термины и определения 2](#_vca0qfc1o41)

[Цель исследования 4](#_kjpltjq70on1)

[**Обзор литературы и предметная область 4**](#_whftrqjr9dwa)

[**Результаты 5**](#_xrdksatog89a)

[Библиотека для симуляции 5](#_804qmyexvx0s)

[Механизмы сосуществования 6](#_xb7ebal9dyrg)

[Механизм Heteromyopia 8](#_k1owj36u6ngj)

[Одномерный случай Heteromyopia 8](#_urrqm4jt5jnz)

[Результаты симуляций 8](#_l2qqargthc4h)

[Результаты численного метода 9](#_bbrd0w6a5qlz)

[Сравнение результатов 10](#_52meoovvtdug)

[Двумерный случай Heteromyopia 10](#_n6myrzkekfay)

[Результаты симуляции 10](#_nt5u8q8j40ky)

[Результаты численного метода 12](#_u9vyifcb4psp)

[Сравнение результатов 12](#_t1cfto2jmfdx)

[Механизм CCTO 12](#_v78skgkzp12d)

[Одномерный случай CCTO 13](#_u68u0kmano5s)

[Результаты симуляций 13](#_t8utbveooe3f)

[Результаты численного метода 14](#_o03jgyvt0ypf)

[Сравнение результатов: 14](#_8lmx9nrrbb3b)

[Двумерный случай CCTO 14](#_xriqx8xderlf)

[Результаты симуляций 14](#_74x8texv5zbi)

[Результаты численного метода 15](#_xuc4488ft8ob)

[Сравнение результатов: 16](#_wk2fhjnvd2z7)

[Применение методов машинного обучения 16](#_994fnsm44ttb)

[**Заключение 18**](#_x3u8jx7yo4ap)

[**Список литературы 18**](#_hb1qiotsrcr9)

## Введение

В современном мире экология и изучение сложных экосистем становятся все более важными задачами. Однако, из-за сложности и разнообразия биологических систем, традиционные аналитические методы не всегда могут полностью объяснить и предсказать их динамику. В таких случаях компьютерные симуляции становятся мощным инструментом для исследования и моделирования экологических процессов.

В истории развития моделирования популяций было предложено множество различных аналитических моделей, как например модель Лотки-Вольтерры, которая описывала популяционную динамику посредством системы дифференциальных уравнений. Однако, с появлением компьютерных технологий и возможности учета индивидуальных особенностей в моделях, стали популярными модели, основанные на индивидах. Эти модели позволяют учесть взаимодействия между индивидами и рассмотреть более реалистичные сценарии. В рамках данного проекта была выбрана модель Дикмана-Лоу для проведения симуляций, так как она представляет собой один из современных подходов в моделировании популяций.

### Основные термины и определения

Стохастическая симуляция - компьютерная симуляция, которая во время каждой генерации использует псевдослучайные значения для определения следующего состояния симулируемой системы.

Стационарное сообщество - сообщество живых организмов, которые в течение жизни не двигаются друг относительно друг друга(в модельной ситуации), как например деревья в лесу или планктон в океане.

IBM модель (Individual Based Models) - модель биологического сообщества развитие которого основано на взаимодействии отдельных индивидов.

Ядро рождения - функция, которая описывает вероятность появления нового индивида на заданном расстоянии от родителя.

Ядро конкуренции - функция, описывающая вероятность некоторого индивида погибнуть от конкуренции с другим индивидом, находящегося от него на заданном расстоянии.

Сосуществование - состояние, при котором два вида сосуществуют в сообществе, ни один из них не доминирует над другим.

Precision - метрика классификации, показывающая долю правильно предсказанных положительных классов от общего числа предсказанных положительных классов.

Recall - метрика классификации, показывающая долю правильно предсказанных положительных классов от общего числа истинных положительных классов

Macro-F1 - метрика классификации, учитывающая precision и recall обоих классов и устойчивая к дисбалансу классов

Датасет - набор данных, используемый для обучения и тестирования моделей машинного обучения

Oversampling - метод балансировки классов путем увеличения числа образцов в меньшем классе

Undersampling - метод балансировки классов путем уменьшения числа образцов в большем классе

Кросс-валидация: метод оценки производительности модели, при котором данные разделяются на обучающий и тестовый наборы несколько раз.

Нейронные сети - модель машинного обучения в которой есть множество объектов, называемых нейронами, которые объединены в слои. Нейроны между соседними слоями в этой модели передают друг другу импульсы.

Наивный байесовский классификатор - вероятностный классификатор, основанный на принципе независимости признаков

Решающие деревья- модель машинного обучения, основанная на том, что строится дерево, в котором у каждой вершины не являющейся листом есть вопрос ответ на который показывает куда дальше двигаться по дереву

Градиентный бустинг - метод машинного обучения, который оптимизированным образом складывает небольшие решающие деревья

Ансамбли деревьев - метод машинного обучения, основанный на усреднении большого количества больших разрешающих деревьев.

Логистическая регрессия - модель машинного обучения, которая применяется для бинарной классификации и основана на линейной регрессии

Метод опорных векторов(SVM) - алгоритм машинного обучения, который разделяет гиперплоскостью пространство, причём при помощи специально подобранных ядер можно регулировать форму гиперплоскости.

### Цель исследования

Целью данного проекта является разработка и проведение компьютерных симуляций, основанных на моделировании стационарных сообществ. Основное внимание уделяется механизмам сосуществования двух видов на участках прямой и плоскости. Основные задачи включают участие в разработке специализированного пакета для проведения симуляций, проведение большого числа симуляций и обработку полученных данных, получение данных численных методов, сравнение результатов численных методов и симуляций. Также задачей является попытка применить методы машинного обучения для обнаружения ситуации сосуществования.

## Обзор литературы и предметная область

Модель Дикмана-Лоу основана на стохастических процессах и позволяет учитывать случайные факторы в динамике популяций. Это позволяет проводить большое количество симуляций и получать разнообразные результаты. Выбор этой модели обусловлен ее способностью учесть различные факторы, влияющие на стационарные сообщества, и пространственную структуру сообщества.

В модели Дикмана-Лоу сообщество рассматривается как набор точек на прямой, плоскости или в трехмерном пространстве (тогда соответствующая модель будет называться одномерной, двумерной или трёхмерной), причём в каждой такой точке находится один индивид. В многовидовом случае каждому индивиду соответствует свой биологический вид, причём может быть так, что все индивиды принадлежат одному и тому же виду. Каждому виду соответствуют свои параметры, в том числе темп размножения, темп смертности и ядро рождения. Индивид порождает потомка того же вида как и он сам, причём для простоты считается, что индивиды сразу рождаются взрослыми.

Особенностью модели является взаимодействие индивидов, поэтому также рассматриваются такие параметры как темп и ядро конкуренции при действии индивида одного вида на индивида другого вида (возможно, что эти два вида совпадают).

Как в модели, так и в симуляции и численных методах существуют следующие обозначения:

|  | Темп рождения i-го вида. В каждый момент времени вероятность того, что индивид i-го вида даст потомство пропорциональна этому коэффициенту |
| --- | --- |
|  | Темп смертности i-го вида от причин не связанных с конкуренцией с другими индивидами. В каждый момент времени вероятность того, что индивид i-го вида умрёт естественной смертью пропорциональна этому коэффициенту |
|  | Сила конкуренции i-го вида по отношению к j-ому виду. В каждый момент времени вероятность того, что индивид i-го вида убъёт индивид j-го вида пропорциональна этому коэффиценту |
|  | Ядро рождаемости i-го индивида. Задаёт пространственное распределение возможностей появления нового индивида i-го вида |
|  | Ядро конкуренции i-го вида по отношению к j-ому. Задаёт пространственное распределение возможностей нападения индивида i-го вида на индивида j-го вида |

Модель удовлетворяет бесконечной систему интегро-дифференциальных уравнений, о чём можно прочитать в [3]. Но для численного решения приходится как-то ограничивать эту систему, для этого используя так называемые “замыкания”, благодаря которым можно искать приближенные решения исходной бесконечной системы

За всеми подробностями обращаться в [3], [4], [8].

## Результаты

Основной темой проекта было исследование механизмов сосуществования двухвидовой модели. Механизм сосуществования - это совокупность наборов параметров симуляции при которых ни один из видов не вымирает. Существует два наиболее известных механизма сосуществования, которые имеют смысл в модели Дикмана-Лоу: это Heteromyopia и Competition-Colonization Trade-Off (далее CCTO). Их описание следует за описанием библиотеки.

Все симуляции и результаты данного проекта расположены в репозитории[7], в котором в README указана структура репозитория.

### Библиотека для симуляции

Для проведения симуляции в высоких размерностях требовалась соответствующая библиотека[2], она написана на языке C++ и в качестве интерфейса для работы с ней используется язык R. Но последняя версия для проведения многовидовых симуляций не обновлялась последние 3 года, что стало некоторым препятствием для работы с нею

При попытках проводить симуляции постоянно появлялись ошибки и странное поведение. Как оказалось странное поведение объяснялось тем, что в библиотеке были проблемы с учётом параметра рождаемости, что было исправлено. Ошибки же сводились к тому, что программа неправильно учитывала влияние особей друг на друга. Из-за этого возможна была ситуация, когда уже мёртвый индивид мог выбираться программой для того, чтобы умереть вновь, что приводило к ошибке. В свою очередь, такое поведение появлялось потому при своём появлении индивид мог не влиять на другого индивида, но когда этот другой умирал, то он влил на первого, эта несимметричность приводила к падению программы. Это происходило из-за неправильной работы итератора по чанкам. Ошибка была исправлена, что позволило далее проводить симуляции.

Также был изменён алгоритм генерации равномерного распределения на сфере. В библиотеке содержалась наивная реализация, которая оказалась неверна. Был имплементирован другой известный алгоритм.

### Механизмы сосуществования

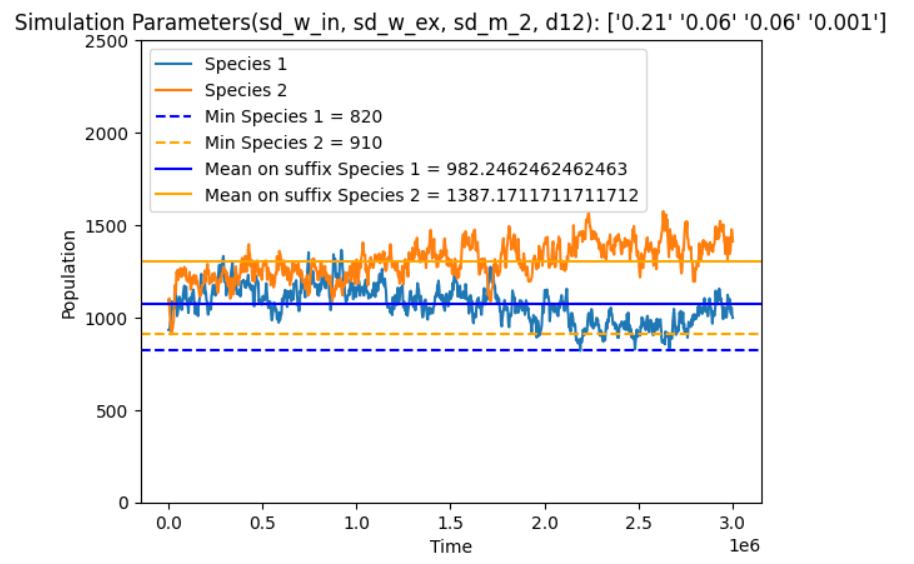
Механизм Heteromyopia[1] относится к соотношению межвидовой и внутривидовой конкуренции. Механизм предполагает, что если межвидовая конкуренция происходит на меньшем расстоянии чем внутривидовая, то виды будут сосуществовать. В модели за это отвечают параметры и .

Механизм CCTO относится к сценарию “хищник-жертва”, когда один вид атакует, а другой убегает. В модели за этот механизм отвечают параметры и , первый вид - хищник, второй -жертва. Предполагается, что при правильной настройке параметров будет происходить сосуществование вида-хищника и вида-жертвы.

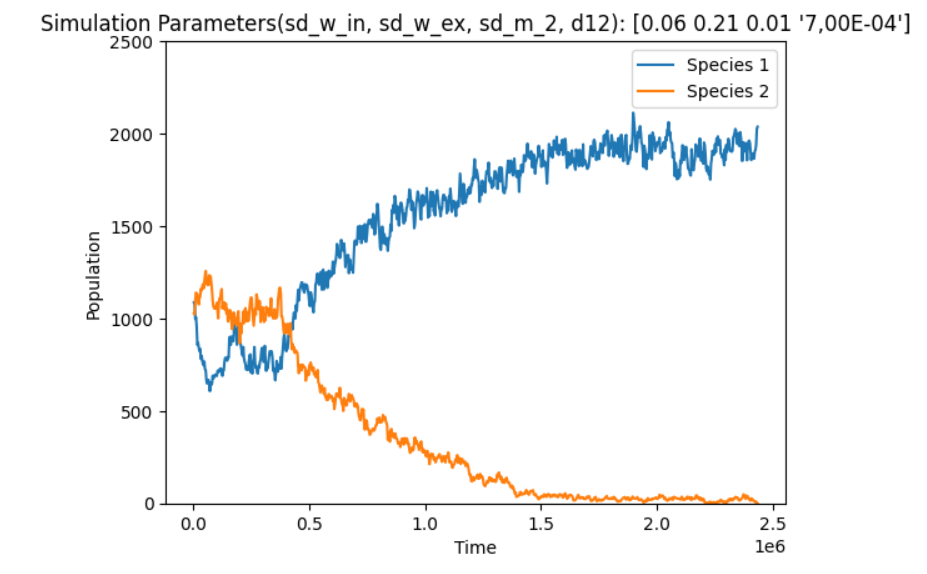
В своей ВКР[4] А.С.Савостьянов рассматривал эти механизмы с точки зрения численных методов, но есть поводы сомневаться в некоторых его результатах в случае высоких размерностей. В этой работе рассматриваются эти механизмы с точки зрения симуляций и проводятся сравнения с новыми результатами численных методов. Для численных методов будет использовано то же замыкание которое использовал Савостьянов в своей работе.

Далее в качестве базовых параметров, которые не участвуют в механизмах сосуществования были взяты параметры из этой статьи[3].

Ниже даны некоторые примеры того как могут вести себя численности видов при сосуществовании и при доминировании одного из видов:



Яркий пример сосуществования, который реализуется при = 0.21, = 0.06, = 0.06, = 0.001.



Случай вымирания вида-жертвы, который реализуется при = 0.06, = 0.21, = 0.01, = 0.007.

Были проведены симуляции для одномерной и двумерной двухвидовой модели в двух сценариях Heteromyopia и CCTO. Они проводились в пространстве параметров , , ,. Результаты симуляции усреднялись, чтобы получить более объективную картину.

### Механизм Heteromyopia

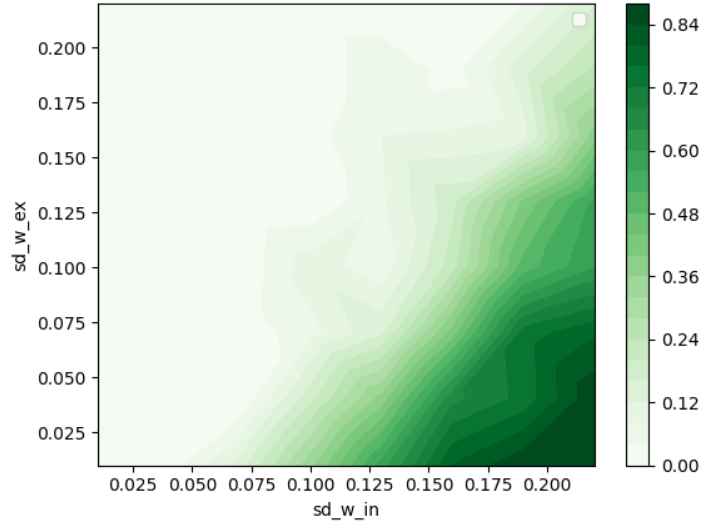
#### Одномерный случай Heteromyopia

В сценарии механизма Heteromyopia существует гипотеза[1], что сосуществование происходит, если радиус ядра межвидовой конкуренции меньше чем радиус ядра внутривидовой конкуренции. Для симуляции этого механизма были зафиксированы = 0.06, = 0.001, первый вид стартовал со 100 особей, второй с 200, параметры были взяты из статьи[3]

##### Результаты симуляций

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

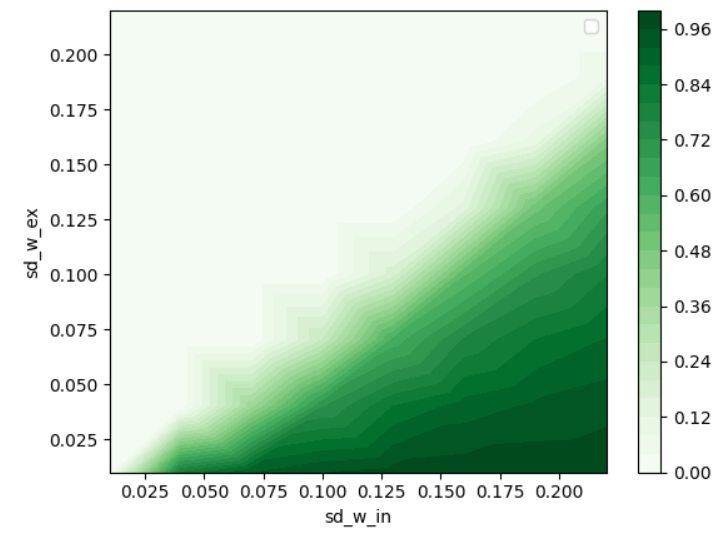
Во всех случаях второй вид выживает, поэтому степень сосуществования можно рассматривать здесь как отношение численности первого вида ко второму виду



##### Результаты численного метода

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

Отношение численности первого вида к численности второго вида



##### Сравнение результатов

| Характеристика | Коэффициент корреляции результатов симуляции и численного метода |
| --- | --- |
| Численность первого вида | 0.84566681 |
| Численность второго вида | 0.85272098 |
| Отношение численности первого вида к численности второго вида | 0.85829182 |

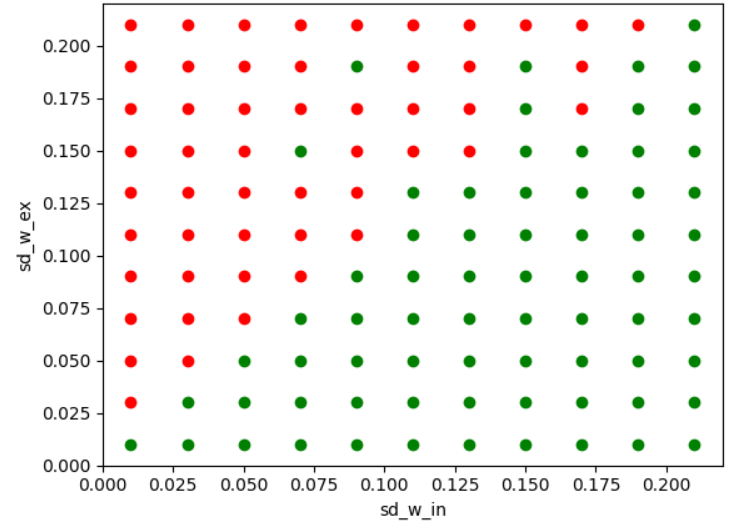
Можно сделать вывод, что в данном случае результаты симуляции и численного метода довольно схожи, также отлична видна предполагаемая линейная граница между ситуациями сосуществования и не сосуществования

#### Двумерный случай Heteromyopia

##### Результаты симуляции

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

График вымирания



Зелёным отмечены ситуации сосуществования, красным не сосуществования. Здесь чётко заметна линейная связь между параметрами и выживаемостью.

##### Результаты численного метода

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

##### Сравнение результатов

Очевидна несостыковка, что означает, что используемое замыкание не применимо в данной ситуации и по рассмотрению результатов численного метода здесь нельзя сделать выводов о результатах симуляции.

### Механизм CCTO

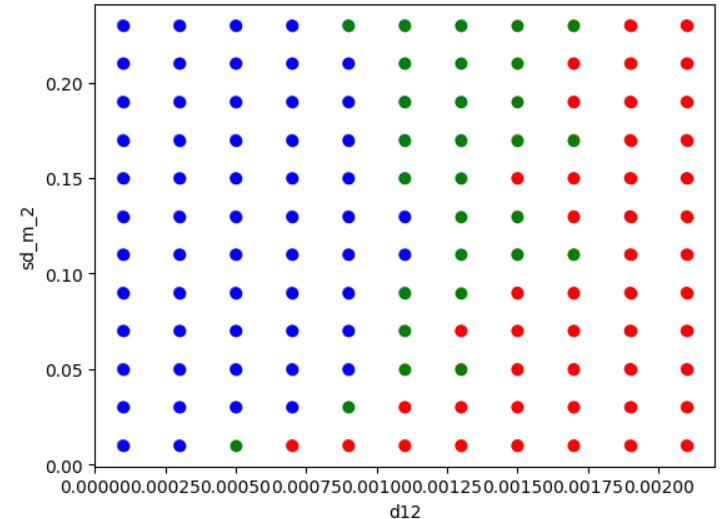
Для симуляции этого механизма были зафиксированы = 0.04, первый вид стартовал со 100 особей, второй с 200, параметры были взяты из статьи[3]

#### Одномерный случай CCTO

##### Результаты симуляций

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

График сосуществования:



Зелёный цвет означает сосуществование, синий цвет означает доминирование вида “жертвы”, красный цвет означает доминирование вида “хищника”

##### Результаты численного метода

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

##### Сравнение результатов:

| Характеристика | Коэффициент корреляции результатов симуляции и численного метода |
| --- | --- |
| Численность первого вида | 0.22647384 |
| Численность второго вида | 0.40720084 |

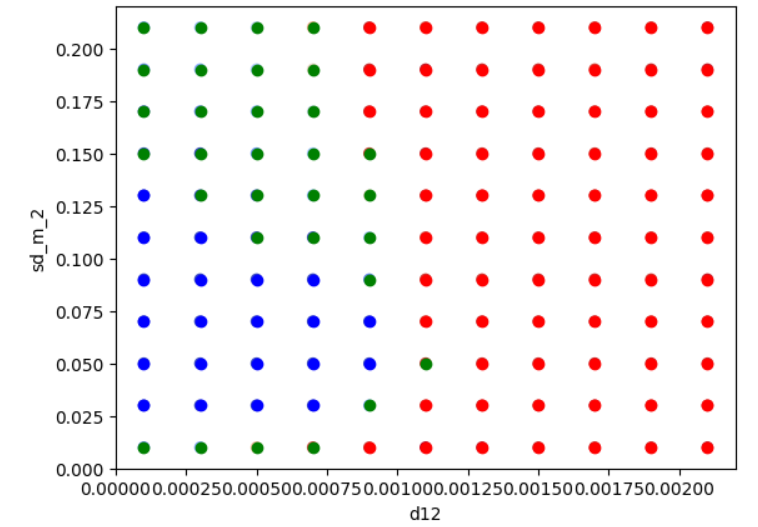
Результаты имеют лишь отдаленное сходство, можно сказать, что между ними нет соответствия, что ставит под вопрос применимость используемого замыкания в данном случае.

#### Двумерный случай CCTO

##### Результаты симуляций

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

График сосуществования:



Зелёный цвет означает сосуществование, синий цвет означает доминирование вида “жертвы”, красный цвет означает доминирование вида “хищника”

##### Результаты численного метода

|  |  |
| --- | --- |
| Численность первого вида | Численность второго вида |

##### Сравнение результатов:

##### 

| Характеристика | Коэффициент корреляции результатов симуляции и численного метода |
| --- | --- |
| Численность первого вида | 0.38422526 |
| Численность второго вида | 0.01300331 |

Хотя графики для первого вида слегка похожи, для второго вида они, можно сказать, почти “ортогональны”. Это показывает, что в данном случае численные методы тоже показывают неправильный результат

### Применение методов машинного обучения

Был составлен датасет из более чем 16000 разнообразных симуляций[7]. Рассматривалась задача бинарной классификации, где все симуляции делились на два класса: сосуществование и не сосуществование. Для задачи классификации рассматривались следующие модели: логистическая регрессия, метод опорных векторов, решающие деревья, ансамбли деревьев, градиентный бустинг, наивный байесовский классификатор и небольшие нейронные сети. Для каждой модели гиперпараметры подбирались методом кросс-валидации. В качестве метрики классификации была выбрана macro-F1. Такой выбор был сделан из-за того, что сама ситуация сосуществования является редкой из-за чего появился дисбаланс классов в датасете. Метрика macro-F1 учитывает precision и recall обоих классов и устойчива к их дисбалансу.

Рейтинговая таблица моделей по показателю macro-F1 на тестовой выборке:

| Модель | Гиперпараметры и замечания | macro-F1 |
| --- | --- | --- |
| Метод опорных векторов | -Реализация из библиотеки sklearn -Параметр C равен 10000  -Параметр “kernel” равен “rbf”  -Параметр “gamma” равен “scale”  -Использовалось радиальное ядро | 0.92 |
| Градиентный бустинг | -Реализация из библиотеки sklearn  -Параметр “learning\_rate” равен 0.3  -Параметр “max\_depth” равен 3  -Параметр “min\_samples\_leaf” равен 1  -Параметр “min\_samples\_splite” равен 2  -Параметр “n\_estimatorы” равен 500 | 0.91 |
| Случайный лес | -Реализация из библиотеки sklearn -Количество деревьев равно 500 -Максимальная глубина ограничена 20 -Параметр “min\_samples\_leaf” равен 2 -Параметр “min\_samples\_split” равен 5  -Использовался критерий Джини | 0.9 |
| kNN | -Рассматриваются 17 соседей  -Расстояние - манхэттенская метрика | 0.89 |
| Решающее дерево | Реализация из библиотеки sklearn -Параметр “min\_samples\_leaf” равен 1  -Параметр “min\_samples\_split” равен 2  -Максимальная глубина не ограничена -Использовался критерий энтропии | 0.88 |
| Небольшие нейронные сети | -Реализация из библиотеки Keras  -Полносвязная сеть  -Два скрытых слоя по 128 нейронов -Функция активации - ReLU -Оптимизатор rmsprop  -Параметр “epochs” равен 1000 | 0.85 |
| Логистическая регрессия | -Параметр “С” равен 10  -Параметр “penalty” равен “l1”  -Параметр “solver” равен “saga”  -Параметр “learning\_rate” равен 0.3 | 0.82 |
| Наивный байесовский классификатор | -Реализация из библиотеки sklearn -Параметр “var\_smoothing” равен 1e-07 | 0.76 |

Методы машинного обучения в целом показали неплохой результат. Неожиданно хорошо показали себя метод опорных векторов, kNN и обычное решающее дерево, можно предположить, что это связано с тем, что границы между зонами сосуществования и не сосуществования имеют простой характер.

Были попытки преодолеть дисбаланс классов с помощью oversampling и undersampling, но при их использовании результаты моделей получались хуже.

Подробнее с тем, что означают параметры можно ознакомиться в документации [5] и [6].

## Заключение

Проведенное исследование позволило глубже понять динамику популяций в экологических системах с использованием компьютерных симуляций на основе модели Дикмана-Лоу. В процессе исследования были применены разнообразные методы анализа, включая статистические модели, машинное обучение и численные методы, что позволило получить более глубокое понимание рассматриваемых экологических процессов. Сделаны выводы о применимости используемого замыкания к механизмам сосуществования. Можно сказать, что в целом при его использовании численные методы показывают неадекватные результаты. За исключением случая 1-мерной Heteromyopia. Поэтому стоит рассматривать и другие замыкания и тестировать их.

Несмотря на достигнутые результаты, следует отметить, что данная работа имеет определенные ограничения, такие как ограниченный объем данных и упрощенные модели, которые могут потребовать дальнейшего углубленного исследования.

## Список литературы

1) David J. Murrell, Richard Law “Heteromyopia and the spatial coexistence of similar competitors”. Ecology letters, 2003

2) Rcpp interface for c++ based poisson simulator. url: https://github.com/YegorGalkin/RcppSim (д. о.: 20.03.2024).

3) А.А.Никитин Е.Г.Галкин В.К. Зеленков. “Компьютерные симуляции и численные методы в двухвидовой модели пространственных сообществ”. International Journal of Open Information Technologies (2019)

4) А.С. Савостьянов: Выпускная квалификационная работа бакалавра на тему “Механизмы сосуществования стационарных биологических сообществ в пространствах разных размерностей”, НИУ ВШЭ, ФКН, ОП “Прикладная математика и информатика”, Москва 2017

5) Scikit-learn: Machine Learning in Python. url: https://scikit-learn.org/stable/index.html (д. о.: 20.03.2024).

6) Keras: The Python Deep Learning API. url: https://keras.io/ (д. о.: 20.03.2024).

7) Наговицын Александр. Scripts and results for Course Work. url: https://github.com/ocelot335/BioSimulations (д. о. 25.03.2024).

8) Richard Law, David J. Murrell, Ulf Dieckmann. On moment closures for population dynamics in continuous space. Journal of Theoretical Biology 229, pages 421–432, 2004