

Классификация траекторий динамических систем с помощью физически-информированных нейросетей

А. А. Терентьев, В. В. Стрижов

terentev.aa@phystech.edu; panchenko.sk@phystech.edu

В работе рассмотрена задача классификации многомерных траекторий физических систем. Классические методы для классификации траекторий как многомерных временных рядов используют предположения о статистических свойствах траекторий, но не используют априорную информацию, что временные ряды являются траекториями динамических систем, подчиняющихся физическим законам и неустойчивы к нефизическим изменениям траекторий. Для учета таких изменений требуется дополнительно обучать модели на аугментированной выборке. Целью работы является предложить метод для классификации многомерных временных рядов, использующие информацию о их физической природе. Для этого предлагается классифицировать не сами траектории, а динамические системы, заданные лагранжианами, соответствующие этим траекториям. Лагранжиан восстанавливается с помощью физически-информированных лагранжевых нейронных сетей. Предложена норма, которая вводится на пространстве лагранжианов. Она используется как метрика для метрических методов классификации. В основном для многомерных классификаций используются сверточные нейронные сети, в предположении, что существуют статистические связи, между ближайшими точками временного ряда. Данные методы хорошо себя показывают во многих задачах, где эти связи являются преобладающими. Но для физических систем такие методы не подходят, и предлагается использовать знания о физических связях систем, для выбранных Лагранжевых нейронных сетей этим знанием является закон сохранения энергии.

Ключевые слова: *физическая система; лагранжиан; лагранжева нейронная сеть; гипотеза компактности*

1 Введение

Физически-информированные нейронные сети (PINN) стремятся объединить преимущества двух подходов к анализу, чисто статистического, основанного на большом количестве данных, когда ничего не известно о физике системы и чисто физического, когда все известно о физике системы, а данные практически не используются. PINN использует знания достаточно общих законов физики, в основном законов сохранений, при этом используя данные как основу для предсказаний. Архитектура подобных систем подробно рассматривается в работе [1]. В последнее время повысился интерес к таким системам,

т.к. физические законы находят отражения в различных областях, при этом основанные на этих законах PINN часто лишены недостатков сетей, которые имеют в своей основе только статистический подход. На основе PINN существуют нейронные сети, учитывающую гамильтонову [15], лагранжеву динамику [8], в том числе гамильтонову механику с бесконечномерным пространством измерений [16], моделирующую квантовую физику. Как показывает работа, лагранжевы сети также позволяют моделировать микросистемы с помощью графовой сети. При этом существуют модификации LNN моделирующие макросистемы, такие как жидкости с помощью многочастичного приближения. PINN и LNN также находят свое отражение в задаче восстановления звука [14], достигающие лучшего результата, за счет учета волнового уравнения, которому подчиняется звук. При этом как показывает работа есть общий подход создания PINN с помощью теории динамических систем, в основе которых лежат интегралы движения.

Работа [12] исследует проблему создания адаптивных гамильтоновых нейронных сетей, которые адаптируются к нелинейному изменению параметризованной системы, и определяющие этот параметр. Этот параметр ответствен за хаотичность системы, таким образом выступающей в качестве меры хаоса и отделяющей хаотические системы от исходной. Недостатком такой классификации является то, что нужно заранее знать закон по которому эти системы отделяются. Согласно работе [13] большинство NODE неустойчивы к изменению начальных данных, ошибка растет экспоненциально по времени. Представленная архитектура CH-NODEs этих недостатков и показывает лучшие результаты по сравнению с ResNet в задачах классификации рисунка. В то же время более общая задача классификации траекторий систем, подчиняющихся лагранжевой динамики, остается неисследованной.

Задача классификация многомерных временных рядов (MTCS) является намного менее изученной задачей по сравнению с одномерным случаем [5]. Несмотря на это данная задача имеет применения в множестве областей: в сфере безопасности, где требуется детектирование критических событий на основе данных с нескольких датчиков [6], определении вида активности людей по данным с акселерометров и гироскопов [7]. Основными методами глубокого обучения для являются методы разработанные для одномерного случая AlexNet, ResNet, InceptionTime, TapNet, основанные на сетях для восстановления траекторий [5]. Из-за того, что MTCS является достаточно широким классом задач методы для его решения практически не содержат в своей архитектуре знаний о задаче, что соответственно требует большого числа данных для той же обобщающей способности модели.

В данной работе предлагается рассмотреть более узкую задачу классификации траекторий динамических систем. Предлагается использовать Лагранжевы нейронные сети (LNN) для векторизации траекторий. Каждой траектории сопоставляется лагранжиан системы, параметризованный LNN. Вопрос сопоставления лагранжиана системы по траектории рассматривается в работе [8].

Лагранжева сеть восстанавливает траектории системы, для которых выполняется закон сохранения энергии и уравнения Эйлера-Лагранжа. LNN позволяет использовать любой независимый набор координат для задания состояния системы. При этом в лагранжевом формализме динамика системы полностью задается функцией лагранжа. Таким образом каждой траектории ставится в соответствии динамическая система, из которой траектория получена, а каждая динамическая система задается по определению лагранжианом.

Для задания состояния системы в лагранжевой механике используются обобщенные координаты. Выбирается минимальное число независимых переменных (обобщённых координат), необходимое для полного описания состояния механической системы.

Лагранжианом системы называется любая непрерывно дважды дифференцируемая функция, интегрируемая в квадрате, для которой уравнения Эйлера-Лагранжа решаются однозначно.

Уравнение Эйлера-Лагранжа имеет вид

$$\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}} = Q_x = 0. \quad (1)$$

Для того чтобы оно решалось однозначно необходимо и достаточно

$$\det \left\| \frac{\partial^2 L}{\partial \dot{q}_j \partial \dot{q}_k} \right\|_{j,k=1}^n \neq 0. \quad (2)$$

Эквивалентное физическое определение лагранжианом как разность между кинетической T и потенциальной энергией V :

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q}, \dot{\mathbf{q}}), \quad (3)$$

где $\mathbf{q}(t)$ — обобщенные координаты системы, $\dot{\mathbf{q}}(t)$ — обобщенные ускорения системы, L — лагранжиан системы, далее для краткости мы опускаем t .

Динамика системы описывается системой уравнений Эйлера-Лагранжа:

$$\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}} = Q_x = 0. \quad (4)$$

Для лагранжианов физических систем утверждается гипотеза компактности: лагранжианы, отвечающие различным движениям, лежат в различных компактных и отделимых подмножествах-классах пространства моделируемых функций. На основе этого делается вывод, что векторы параметров функций, соответствующих различным классам, также компактны и отделимы в общем пространстве параметров.

Классификация проведена на наборе данных РАМАР2. Набор данных содержит траектории движений человека во время обычных дневных активностей: готовка еды, пробежка, поход до магазина и другие.

2 Постановка задачи регрессии для временного ряда и Лагранжевы нейронные сети.

Каждой траектории ставится в соответствие лагранжиан динамической системы L . Лагранжиан задает динамическую систему, для которой траектория получена, т.к. при подстановке L в уравнение Эйлера-Лагранжа (4) получается функция, сопоставляющая скорости и координатам точки ее ускорение. Решив систему ОДУ, полученную при подстановке L в уравнение Эйлера-Лагранжа (4), при начальных условиях равным начальной точки траектории, траектория полностью восстанавливается. Таким образом мы векторизуем каждую траекторию функцией L из пространства L_2 и отождествляем траекторию и динамическую систему, заданную функцией L которую в дальнейшем подаем на вход классификатору.

Задача восстановления лагранжиана L формулируется как задача регрессии.

Определение 1. *Траекторией размерности r и длиной T назовем $X = [X_1, X_2, \dots, X_T]$ такой, что $\mathbf{X}_i = (\mathbf{q}_i, \dot{\mathbf{q}}_i)$, $\dot{\mathbf{q}}_i \in \mathbb{R}^r$ – скорость в i -ый момент времени, $\mathbf{q}_i \in \mathbb{R}^r$ – координата в i -ый момент времени*

Дана выборка $D = \{(X_i, \dot{X}_i)\}_{i=1}^N$, где X – траектории размерности r и длиной T , $\mathbf{y}_i = \dot{X}_i = \dot{\mathbf{q}}_i \in \mathbb{R}^r$ – ускорение в i -ый момент времени.

Требуется найти функцию $\hat{\mathbf{y}} = f(\mathbf{X} = (\mathbf{q}, \dot{\mathbf{q}})|D)$, где D – данная обучающая выборка, $\hat{Y} = \hat{X} = \{\hat{\mathbf{y}}_i = \hat{\dot{\mathbf{q}}}_i\}_{i=1}^N$ – предсказанная динамика траектории

При этом функция f в общем случае ищется в семействе $f(\mathbf{x}|\mathbf{w})$ лагранжевых нейронных сетей структура которых рассматривается далее, где \mathbf{w} – это параметры нейронной сети.

В качестве функции потерь и критерия берется среднеквадратичное отклонение значения предсказанного ускорения $\hat{\mathbf{y}}_i$ на выборке из точек $\mathbf{X}_i = (\mathbf{q}_i, \dot{\mathbf{q}}_i)$ траектории X

$$\mathcal{L}(\mathbf{w}) = \mathcal{L}(\mathbf{w}|D) = \frac{1}{T} \sum_{i=1}^T \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2^2. \quad (5)$$

По параметрам нейронной сети ставится оптимизационная задача

$$\mathbf{w}_j^* = \arg \min_{\mathbf{w} \in \mathbb{W}} (\mathcal{L}(\mathbf{w})). \quad (6)$$

Рассмотрим алгоритм моделирования динамики системы в лагранжевой динамике:

1. Найти аналитические выражения для кинетической (T) и потенциальной энергии (V).
2. Получить лагранжиан

$$\mathcal{L} = T - V.$$

3. Применить ограничение Эйлера-Лагранжа

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}} = \frac{\partial L}{\partial \mathbf{q}}.$$

4. Решить получившуюся систему дифференциальных уравнений.

В работе [2] соответственно предложено в нейронную сеть

$$f : \mathbf{X} = (\mathbf{q}, \dot{\mathbf{q}}) \rightarrow \mathbf{Y}. \quad (7)$$

добавить априорные знания о физике системы, моделируя лагранжиан:

$$f : \mathbf{X} = (\mathbf{q}, \dot{\mathbf{q}}) \rightarrow \mathbf{L}. \quad (8)$$

Ключевой идеей является параметризация нейронной сетью лагранжиана L , получение выражения ограничения Эйлера-Лагранжа и обратное распространение ошибки через полученные ограничения

$$\ddot{\mathbf{q}} = (\nabla_{\dot{\mathbf{q}}} L)^{-1} [\nabla_{\mathbf{q}} L - (\nabla_{\dot{\mathbf{q}}} L) \dot{\mathbf{q}}]. \quad (9)$$

Таким образом получаем схему нейронной сети представленную на Рисунке 1.

$$\mathbf{x}_i = (\mathbf{q}_i, \dot{\mathbf{q}}_i) \xrightarrow{\hat{L}: (\mathbf{x}|\mathbf{w}) \rightarrow L_{\mathbf{x}}} L_{\mathbf{x}} \xrightarrow{\nabla L} \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}} = \frac{\partial L}{\partial \mathbf{q}} \xrightarrow{\ddot{\mathbf{q}}_i = g(\mathbf{q}_i, \dot{\mathbf{q}}_i)} \mathbf{y}_i = \ddot{\mathbf{q}}_i$$

Рис. 1 Схема нейронной сети

Рассмотрим структуру сети на Рисунке 1 подробнее. В качестве нейронной сети $\hat{L}: (\mathbf{x}|\mathbf{w}) \rightarrow L_{\mathbf{x}}$ берется полносвязная сеть с тремя слоями и функцией активации SoftPlus, а \hat{L} является оценкой лагранжиана L динамической системы, для которой была получена траектория.

Далее оценка \hat{L} дифференцируется и получается значение $\nabla_{\dot{\mathbf{q}}\dot{\mathbf{q}}} \hat{L}$, $\dot{\mathbf{q}} \hat{L}$, $\nabla_{\dot{\mathbf{q}}\mathbf{q}} \hat{L}$ в точке \mathbf{x} .

Для получения ускорения $\hat{\mathbf{y}}$ полученные значения градиентов подставляется в уравнение Эйлера-Лагранжа (4), которое представляется в виде (9), из которого получается искомая оценка $\hat{\mathbf{y}}$

Для заданных координат $(\mathbf{q}, \dot{\mathbf{q}})$ имеем модель с априорными знаниями о законе сохранения энергии, с помощью которой можем получить динамику параметров $(\dot{\mathbf{q}}, \ddot{\mathbf{q}})$.

3 Постановка задачи классификации траекторий физических систем

Дана выборка $D = \{(X_i, y_i)\}_{i=1}^N$, где X_i – траектории размерности r и длиной T , $y_i \in \overline{1, K}$ – метка i -ой траектории

Требуется найти метод классификации $p(\hat{y}|X, D)$, где $X = \{(X_i)\}_{i=1}^N$ – набор длинны N траекторий размерности r и длиной t , D – данная обучающая выборка, $\hat{y} \in \overline{1, K}^N$ – предсказанные метки классов

Модель классификатора $p(\hat{y}|X, D)$, где $X = \{(X_i)\}_{i=1}^N$ – набор длинны N траекторий размерности r и длиной t , D – данная обучающая выборка, $\hat{y} \in \overline{1, K}^N$ – предсказанные метки классов

Модели сравниваются с помощью метрики Асигасы $= \frac{1}{N} \sum_{i=1}^N [\hat{y}_i = y_i]$, \hat{y}_i – предсказанные моделью метки траектории, \hat{y} – класс, которому принадлежит траектория

Несмотря на такую постановку задачи обычно физически обосновано классифицировать динамические системы, которые в рассматриваем случае отождествляются с лагранжианами L , а не сами реализации систем в виде траекторий.

Поэтому в работе рассматривается несколько другая, но в данной нам интерпретация физически эквивалентная постановка задачи.

Требуется найти метод классификации $p(\hat{y}|X, D) = p(\hat{y}|L, D)p(L|X)$, где $X = \{(X_i)\}_{i=1}^N$ – набор длинны N траекторий размерности r и длиной t , D – данная обучающая выборка, $\hat{y} \in \overline{1, K}^N$ – предсказанные метки классов, $p(L|X)$ – метод сопоставления траектории X лагранжиану L

В работе в качестве метода $p(L|X) = [L = \hat{L}]$ рассматривается точечная оценка траектории X параметризованной оценкой лагранжиана \hat{L} лагранжевой нейронной сетью из рассмотренной задачи регрессии для временного ряда. Но такая постановка дает возможность использовать вероятностную оценку для лагранжиана L – $p(L|X)$.

4 Норма в пространстве лагранжианов и её аппроксимация

Мы считаем, что для лагранжианов траекторий верна гипотеза компактности, т. е., что лагранжианы лежат в компактно локализованных подмножествах. В основе этого предположения лежит то, что в основе различных движений человека лежат различные группы мышц. А в схожих соответственно схожие. Поэтому разные по природе движения будут образовывать разные механические системы и наоборот. Отсюда естественно вытекает гипотеза.

Осталось эту гипотезу проверить. Для этого необходимо уточнить термины, упомянутые в первом абзаце.

Пусть \mathbb{L} — функциональное линейное пространство лагранжианов всевозможных систем:

$$g : \mathbf{x} \rightarrow L, \quad g \in \mathbb{L},$$

где $\mathbf{x} \in \mathbb{R}^{2 \times r}$ — вектор обобщенных координат и скоростей, $L \in \mathbb{R}$ — значение лагранжиана системы. Пусть также имеется параметрическое семейство функций, аппроксимирующих лагранжиан, описанное в разделе 1,

$$\mathbb{L}_\varepsilon = \{\hat{L}^{(\varepsilon)} : (\mathbf{x}|\mathbf{w}) \rightarrow \hat{L}(x) \mid \mathbf{x} \in \mathbb{R}^{2 \times r}, \hat{L}(x) \in \mathbb{R}, \mathbf{w} \in \mathbb{W}\}, \quad \mathbb{L}_\varepsilon \subset \mathbb{L},$$

где \mathbf{w} — параметры функции из линейного пространства всевозможных значений параметров \mathbb{W} , ε — точность, с которой аппроксимируется лагранжианы.

Для построения нормы получим систему уравнений для вычисления ускорений.

Лагранжев формализм моделирует физическую систему с координатами $x_t = (q, \dot{q})$, которая начинается в состоянии x_0 и заканчивается в другом состоянии x_1 . Определяется функционал, который называется действием

$$S = \int_{t_0}^{t_1} L dt, \tag{10}$$

где L — лагранжиан системы. Действие определяет путь, по которому координаты x_t пройдут из x_0 в x_1 в промежуток времени от t_0 до t_1 . Путь минимизирует действие S , т.е. $\delta S = 0$. Это приводит к уравнению Эйлера-Лагранжа, определяющему динамику системы

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}} = \frac{\partial L}{\partial \mathbf{q}}. \tag{11}$$

Ускорение каждой компоненты системы $\ddot{\mathbf{q}}$ может быть напрямую получено из правила цепочки: Распишем производную по времени через производные по координатам:

$$\left(\frac{\partial}{\partial \mathbf{q}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \right) \frac{d\mathbf{q}}{dt} + \left(\frac{\partial}{\partial \dot{\mathbf{q}}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \right) \frac{d\dot{\mathbf{q}}}{dt} = \frac{\partial L}{\partial \mathbf{q}}.$$

Перепишем производные по времени для наглядности

$$\frac{\partial}{\partial \mathbf{q}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}} + \frac{\partial}{\partial \dot{\mathbf{q}}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \ddot{\mathbf{q}} = \frac{\partial L}{\partial \mathbf{q}}.$$

Перенесем слагаемые так, чтобы слева остались только неизвестные

$$\frac{\partial}{\partial \dot{\mathbf{q}}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \ddot{\mathbf{q}} = \frac{\partial L}{\partial \mathbf{q}} - \frac{\partial}{\partial \mathbf{q}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}}.$$

Выразим неизвестную переменную разделив обе части на одно выражение

$$\ddot{\mathbf{q}} = \left(\frac{\partial}{\partial \dot{\mathbf{q}}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \right)^{-1} \left[\frac{\partial L}{\partial \mathbf{q}} - \frac{\partial}{\partial \mathbf{q}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}} \right].$$

Перепишем частые производные в другой нотации

$$\ddot{\mathbf{q}} = (\nabla_{\dot{\mathbf{q}}\dot{\mathbf{q}}} L)^{-1} [\nabla_{\mathbf{q}} L - (\nabla_{\dot{\mathbf{q}}\mathbf{q}} L) \dot{\mathbf{q}}], \quad (12)$$

где гессиан

$$H = \nabla_{\dot{\mathbf{q}}\dot{\mathbf{q}}} L = \frac{\partial^2 L}{\partial \dot{\mathbf{q}}^T \partial \dot{\mathbf{q}}}. \quad (13)$$

Замечание 1. $\det H \neq 0$ в любой точке, а значит он положительно определен. Данное условие необходимо, т. к. иначе существует малая вариация, при которой состояние системы не меняется. Следовательно, имеем не вырожденную систему линейных уравнений относительно ускорений.

$$H \ddot{\mathbf{q}} = \frac{\partial L}{\partial \mathbf{q}} - \frac{\partial}{\partial \mathbf{q}^T} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}}, \quad (14)$$

Лемма 1. Если лагранжианы заданы на компакте, то существуют неотрицательные числа a, b такие, что $\det H \geq a$, а собственные значения матрицы H не меньше b

Доказательство. Предположим, что это не так. H является непрерывной функцией, т.к. ее значения являются вторыми производными L , следовательно, $\det H$ и собственные значения также непрерывны, а значит достигают своего минимума и максимума. Получили противоречия, т.к. матрица невырожденная во всех точках. \square

Замечание 2. Если дополнительно предположить, что система уравнений не является вырожденной в пределе, то утверждение выше верно и для всего пространства

Т.к. уравнения инвариантны относительно умножения Лагранжиана на константу, то мы можем выбрать лагранжины для систем так, чтобы собственные значения матрицы H были не меньше 1.

Теперь видно, что на пространстве \mathbb{L} можно задать норму.

Лемма 2. Оператор $A(L) = \frac{\partial L}{\partial \mathbf{q}} - \frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}}$ является линейным

Доказательство.

$$A(L) = \frac{\partial \alpha L_1 + \beta L_2}{\partial \mathbf{q}} - \frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial \alpha L_1 + \beta L_2}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}} = \frac{\partial \alpha L_1}{\partial \mathbf{q}} + \frac{\partial \beta L_2}{\partial \mathbf{q}} - \left(\frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial \alpha L_1}{\partial \dot{\mathbf{q}}} + \frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial \beta L_2}{\partial \dot{\mathbf{q}}} \right) \dot{\mathbf{q}} =$$

$$\alpha \frac{\partial L_1}{\partial \mathbf{q}} + \beta \frac{\partial L_2}{\partial \mathbf{q}} - \alpha \frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial L_1}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}} - \beta \frac{\partial}{\partial \dot{\mathbf{q}}} \frac{\partial L_2}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}} = \alpha A(L) + \beta A(L)$$

□

Теперь рассмотрим \mathbb{L} как пространство L_2 . В нем $\|L\|_L = \|(A(L), H(L))\|_2$ является полунормой, т.к. линейный оператор сохранит абсолютную однородность и неравенство треугольника.

Но множество $\|L\|_L = 0$ также является подпространством L_2 , т.к. $A(L)$, $H(L)$ - линейные операторы, а значит их ядра линейные пространства. Тогда можно ввести факторпространство, в котором отношением эквивалентности будет равенство 0 полунормы $\|L\|_L$.

Рассмотрим обоснованность введённого отношения эквивалентности.

Лемма 3. $\|L\|_L = 0 \Leftrightarrow$ п.в. $\delta \ddot{q} = 0$, где $\delta L, \delta \ddot{q}$ являются вариациями лагранжиана и ускорения.

Доказательство.

$$\|(A(\delta L), H(\delta L))\|_2 = 0 \Leftrightarrow \text{п.в. } A(\delta L) = 0, H(\delta L) = 0$$

$$A(L) = H(L)\ddot{q}$$

$$A(L + \delta L) = H(L + \delta L)(\ddot{q} + \delta \ddot{q})$$

$$A(\delta L) = H(\delta L)(\ddot{q}) + H(\delta L + \delta L)(\delta \ddot{q})$$

$$0 = H(\delta L + \delta L)(\delta \ddot{q}) \Leftrightarrow \delta \ddot{q} = 0$$

□

Замечание 3. Заметим, что в большинстве случаев $\|H(\delta L)\|_2 = 0$, т.к. всегда можно выбрать координаты так, с помощью преобразования Лежандра, чтобы матрица Гессе была единичной, для этого необходимо перейти в канонические координаты, а значит достаточно сравнивать только первую часть нормы $\|A(L)\|_2$

Данную норму вычислить точно в общем случае невозможно, т.к. для этого нужно посчитать интеграл функции на всей числовой прямой, но его можно приблизить

$$\begin{aligned} \|(A(L).H(L))\|_2 &= \sqrt{\int |A(L)(\mathbf{q}, \dot{\mathbf{q}})|^2 + \|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2^2 d\Omega} \approx \\ &\sum_{i_1=1, \dots, i_n=1}^{N, \dots, N} |A(L)(\mathbf{q}, \dot{\mathbf{q}})|^2 + \|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2^2 \frac{\Delta X_1 \cdot \dots \cdot X_n}{N \cdot \dots \cdot N} = \\ C \cdot \sum_{i_1=1, \dots, i_n=1}^{N, \dots, N} \frac{A(L)(\mathbf{q}, \dot{\mathbf{q}})^2 + \|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2^2}{N \cdot \dots \cdot N} &= C \cdot \overline{|A(L)(\mathbf{q}, \dot{\mathbf{q}})|^2 + \|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2^2}. \end{aligned}$$

Т.к. мы решаем задачу классификации, то сопоставим каждому лагранжиану g его класс \mathcal{A}_i . Таким образом мы имеем конечное семейство непересекающихся замкнутых выпуклых множеств \mathcal{A} в пространстве \mathbb{L} лагранжианов, где каждое семейство представляет класс, соответствующий движению человека

Теорема 1. Пусть есть конечное семейство непересекающихся замкнутых выпуклых множеств \mathcal{A} в нормированном пространстве \mathbb{L} , тогда существует $\varepsilon > 0$, что для любого преобразования пространства φ такое, что $\|\varphi(\mathcal{A}_i) - \mathcal{A}_i\| < \varepsilon$ множества из семейства $\varphi\mathcal{A}$ попарно сильно отделимы.

Доказательство. Согласно первой теореме Хана-Банаха об отделимости множества из семейства \mathcal{A} попарно отделимы

$$\langle a, x \rangle_{x \in \mathcal{A}_i} \leq \gamma_1 < \gamma_2 \leq \langle a, y \rangle_{y \in \mathcal{A}_j}.$$

Т.к. функционал a непрерывен, то для любого $\varepsilon > 0$ существует $U(0)$ — окрестность нуля, такая, что

$$\langle a, y \rangle_{y \in U(0)} \leq \varepsilon.$$

Пусть

$$\varepsilon_{ij} = \frac{|\gamma_1 - \gamma_2|}{3},$$

тогда

$$\langle a, x \rangle_{x \in \mathcal{A}_i + U(0)} \leq \gamma_1 + \varepsilon < \gamma_2 + \varepsilon \leq \langle a, y \rangle_{y \in \mathcal{A}_j + U(0)}.$$

Если взять $\varepsilon = \min_{i,j} \varepsilon_{ij}$, то тогда множества \mathcal{A}_i попарно сильно отделимы. □

Это теорема позволяет исследовать свойства лагранжианов на их приближениях, что заметно упрощает задачу.

Исходя из данной теоремы, можно задать достаточное свойство для алгоритма аппроксимации. Должно существовать $\varepsilon > 0$, такое, что для любого $\mathcal{L} \in \mathbb{L}$ было верно $\|\mathcal{L} - \mathcal{L}_\varepsilon\| < \varepsilon$. Иными словами нужно, чтобы расстояние от аппроксимированного лагранжиана до истинного было небольшим. Т.е. требуется чтобы нейросеть не переобучалась и с достаточной точностью приближала лагранжиан. Данное условие выполняется для корректно построенной сети.

Уже из Теоремы 1 следует первый метод классификации лагранжиана. Т.к. в пространстве \mathbb{L} мы ввели норму, то она порождает метрику. Следовательно, можно использовать метрические методы классификации. В данной же работе предлагается использовать для классификации параметры нейронной сети, аппроксимирующей лагранжиан. В этом нам помогает следующая теорема.

Теорема 2. Пусть есть семейство конечное семейство множеств \mathcal{A} в нормированном пространстве \mathbb{L} , $\varepsilon > 0$, и преобразования пространства φ такое, что $\|\varphi(\mathcal{A}_i) - \mathcal{A}_i\| < \varepsilon$ множества из семейства $\varphi\mathcal{A}$ попарно сильно отделимы, а также взаимно-однозначное непрерывное отображение $\psi : \mathbb{L}_\varepsilon \rightarrow \mathbb{R}^n$, обратное отображение непрерывно и выпукло. Тогда множества \mathcal{A}_i попарно сильно отделимы.

Доказательство. Погрузимся в пространство \mathbb{R}^n . В нем $\varphi\psi\mathcal{A}_i$ попарно сильно отделимы. Тогда каждое множество ограничено соответствующими гиперплоскостями $\langle a, x \rangle \leq \gamma_1$
 $x \in \varphi\psi\mathcal{A}_i$

Расширим каждое множество до гиперплоскостей и будем рассматривать уже расширенные. Каждое из таких множеств компактно. Т.к. ψ^{-1} выпукло и непрерывно, то образ расширенных множеств также выпуклый и компактен. Следовательно, расширенные множества попарно сильно отделимы. \square

Теорема дает достаточные условия для того, чтобы для изначальных множеств выполнялась гипотеза о компактности. Но ее условия достаточно сильные. Для нее требуется, чтобы веса были непрерывны, что может не выполняться. Данный метод удобен тем, что можно не использовать много признаков, но он не подходит для доказательства того, что для набора данных не выполняется гипотеза. В случае ее невыполнения лучше использовать методы основанные на метрике, заданной нормой.

5 Функция потерь

Решение уравнения (14) неизменно относительно \ddot{q} , при умножении L на $c > 0$. В таком случае при стремлении L к нулю, ошибка $\delta\ddot{q}$ стремиться к бесконечности. Также это означает, что решение оптимизационной задачи (6) существенно неоднозначно. Для регуляризации задачи накладываются ограничение на найденное решение

$$\lambda(H) > 1. \quad (15)$$

В соответствие с Леммой 1, решение на множестве с ограничением (15) не хуже, чем для задачи без ограничений. В то же время данное ограничение не решает проблему с неоднозначностью задачи. Для этого требуется модифицировать функцию потерь. Добавляется слагаемое равное невязке при решении СЛАУ (14) относительно \ddot{q}

$$\begin{aligned} H\ddot{\mathbf{q}} &= \mathbf{b}, \\ H\hat{\mathbf{q}} &= \hat{\mathbf{b}} = (\nabla_{\dot{\mathbf{q}}\dot{\mathbf{q}}}L)^{-1} [\nabla_q L - (\nabla_{\dot{\mathbf{q}}\mathbf{q}}L) \dot{\mathbf{q}}], \\ \mathcal{L}^{\text{add}}(\mathbf{w}) &= \frac{1}{T} \sum_{i=1}^T \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|_2^2. \end{aligned} \quad (16)$$

В соответствие с Леммой 3 для \hat{L} дающих точную оценку \ddot{q} $\mathcal{L}^{\text{add}} = 0$. При том при выполнении ограничений (15), $\mathcal{L}^{\text{add}} > \mathcal{L}$ и можно заменить функцию потерь на \mathcal{L}

Таким образом получили следующую функцию потерь, с учетом ограничений

$$\mathcal{L} = \mathcal{L}_{\text{add}} + \mathcal{L}_{\text{restriction}}, \quad (17)$$

$$\mathcal{L}_{\text{add}}(\mathbf{w}) = \frac{1}{T} \sum_{i=1}^T \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|_2^2, \quad (18)$$

$$\mathcal{L}_{\text{restriction}}(\mathbf{w}) = +\infty[(H) > 1]. \quad (19)$$

6 Метод классификации

Метод классификации имеет следующий вид $p(\hat{y}|X, D) = p(\hat{y}|L, D)p(L|X)$. Как было сказано в разделе 1 метод $p(L|X)$ ставит в соответствие каждой траектории оценку лагранжева \hat{L} из задачи линейной регрессии для лагарнжевой нейронной сети.

Но просто применить векторный метод классификации L в качестве $p(\hat{y}|L, D)$ невозможно, т.к. L вектор бесконечномерного пространства \mathbb{L} . Для этого пространство L проецируется на конечномерное пространство так, чтобы норма рассматриваемых нами векторов менялась не более чем на заранее заданную величину $\varepsilon > 0$

В качестве нормы на пространстве \mathbb{L} берется норма $\|L\|_L$. Для вычисления нормы используется ее аппроксимация на конечном наборе точек $\|L\|_L \approx \sqrt{|A(L)(\mathbf{q}, \dot{\mathbf{q}})|^2 + \|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2^2}$. Если взять значения $A(L)(\mathbf{q}, \dot{\mathbf{q}})$ и $\|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2$ в качестве значения вектора, то получается проекция пространства \mathbb{L} на конечномерное пространство. Данная проекция и подается на вход классификатору. Таким образом получаем метод классификации представленный на Рисунке 2.

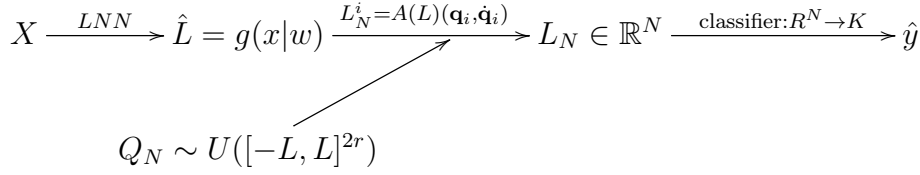


Рис. 2 Схема предложенного классификатора траекторий

Рассмотрим схему на Рисунке 2. В начале на этапе обучения заранее выбирается набор точек $Q_N \sim U([-L, L]^{2r})$, $|Q_N| = N$ в которых сэмплируются значения функций. Для траектории X находится оценка лагранжиана \hat{L} лагранжевой нейронной сетью. Потом для функции \hat{L} ищется значения $A(L)(\mathbf{q}, \dot{\mathbf{q}})$ и $\|H(L)(\mathbf{q}, \dot{\mathbf{q}})\|_2$ в выбранных точках Q_N . Из этих значений собирается вектор L_N . Далее этот вектор L_N подается на вход векторному классификационному методу. Таким образом точки Q_N выступают в роли гиперпараметром, а обучаемой частью модели является классификатор $C : L_N \rightarrow \hat{y}$.

7 Вычислительный эксперимент

Целью эксперимента является классификация многомерных временных рядов, представленных в виде траекторий физических систем. Как подзадачи необходимо подтвердить гипотезу о компактности классов, подтвердить на практике выдвинутую теорию, что пространство лагранжианов можно нормировать предложенной нормой и сэмплировать ее заранее заданным набором точек и с помощью нее построить классификатор.

Исследования проводились на наборе данных Physical Activity Monitoring(PAMAP2) [?]. Набор данных содержит записи с трех наборов гироскопов и акселерометров: закрепленных на запястье преобладающей руки, закрепленных на груди, закрепленных на локте преобладающей руки. Число испытуемых: $M = 9$. Число видов активностей(классов): $K = 24$, каждая активность длилась 2-4 минуты [?], частота сэмплирования 100 Гц.

Из всех датчиков представленных в наборе данных были выбраны гироскопы, как наиболее подходящие для классификации. ходьбы, бега, поездки на велосипеде и неподвижного состояния. Обработка данных состоит в восстановлении пропусков в данных,

связанных с отсутствием сигналов от гироскопов, а также получению значения углов и угловых ускорений из угловых скоростей, полученных от гироскопов.

Часть данных в рядах потеряно из-за проблем в связи. Для восстановления ряда используется сплайн-интерполяция, которая позволяет получить квадратичную точность от размеров сетки. Для восстановления ускорения \ddot{q} , скорости \dot{q} и координат q применяется численное дифференцирование и интегрирование: для дифференцирования используется центральная разность –

$$\ddot{q}_k = \frac{\dot{q}_{k+1} - \dot{q}_{k-1}}{2h}, \quad |E(\ddot{q}_k)| \leq \frac{h^2}{6} q^{(4)}, \quad (20)$$

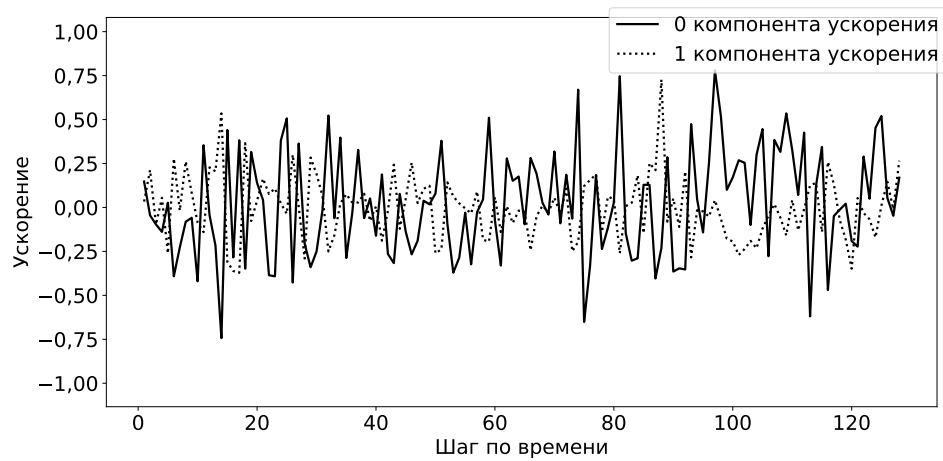
для интегрирования используется формула Кортеса –

$$q_k = \int_0^{t_k} \dot{q}(t) dx \approx \frac{h}{3} \sum_{k=1}^{N-1} (\dot{q}_{k-1} + 4\dot{q}_k + \dot{q}_{k+1}), \quad (21)$$

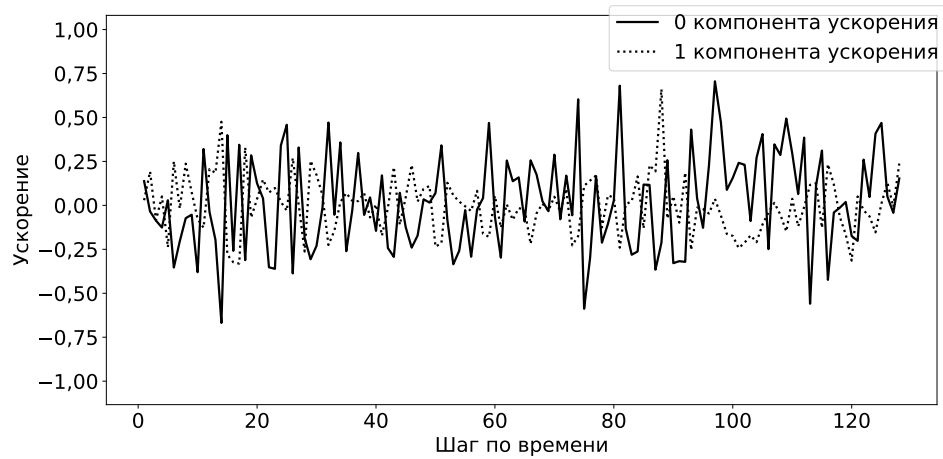
$$|E(q_k)| \leq \frac{(b-a)}{2880} h^4 \max |q^{(4)}(x)|. \quad (22)$$

Полученные траектории передаются как объекты $\mathbf{X}_i = (\mathbf{q}, \dot{\mathbf{q}})(t)_i$ и признаки $\mathbf{Y}_i = \ddot{\mathbf{q}}(t)_i$. Нейросеть, представленная в работе [8], обучается на i -ой траектории отдельно и получается оценка лагранжиана \hat{L}_i , сопоставленный i -ой траектории. Этот лагранжиан, задает динамическую систему и позволяет восстановить ускорение $\ddot{\mathbf{q}}(t)_i$. Таким образом получается набор объектов \hat{L}_i , и набор меток y_i , где y_i - это класс траектории X_i .

В качестве визуализации и оценки восстановления траекторий с необходимой точностью построены предсказания LNN ускорения $\ddot{\mathbf{q}}(t)$ для траекторий $(\mathbf{q}, \dot{\mathbf{q}})(t)$ и сравнены с реальными.



(а)



(б)

Рис. 3 Временной ряд зависимости ускорения от времени для тестовой выборки (а) реальные данные, (б) предсказанные

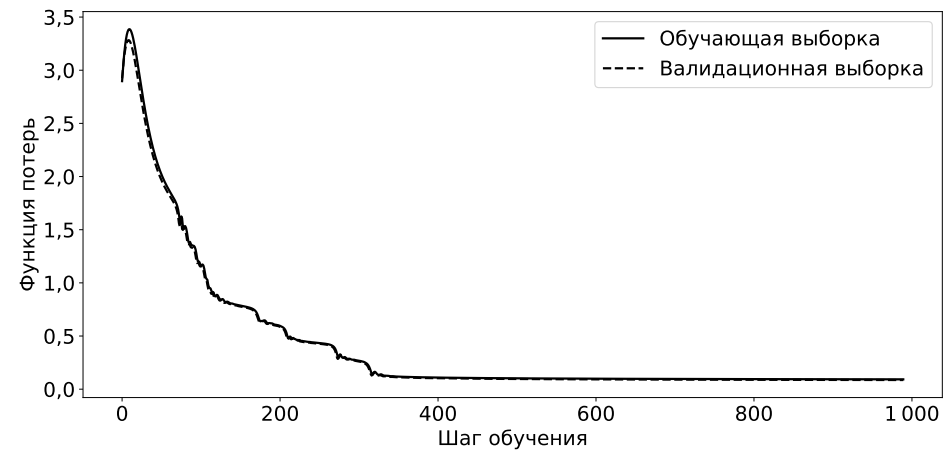


Рис. 4 Графики обучения модели

Рисунки 3, 4 показывают, что траектории в точности восстанавливаются их лагранжианом, причем мы можем предсказывать их на достаточно большом отрезке времени, потери на тестовой и обучающей выборке близки, а траектории сохраняют свой вид. Это означает, что в соответствие с предположением пространство Лагранжианов является скрытым пространством, из которого сгенерированы все траектории.

Таким образом \hat{L}_i векторизует данные X_i для дальнейшей классификации. В замечании к лемме 2 введена норма на пространстве лагранжианов, что представляет его как гильбертово пространство. Окончательно, векторизации берется из замечания 3, в котором пространство лагранжианов проецируется на евклидово пространство \mathbb{R}^n . Векторизация траектории соответствует векторизации их лагранжианов.

В отличие от работы [8] предлагается дополнительно регуляризовать функцию потерь. Т. к. в областях, где собственные числа матрицы $H = (\nabla_{\dot{\mathbf{q}}} L)_{ij} = \frac{\partial^2 L}{\partial \dot{q}_j \partial \dot{q}_i}$ стремятся к 0 решение является неустойчивым, а на траекториях выявляются нефизические колебания. Поэтому вводится дополнительный член $\mathcal{L}_{\text{restriction}} = \sum \alpha * a(\beta(1 - \lambda(H)_i))$, где a некоторая линейная функция активации, лучше всего себя показала $a = \text{SiLU} = x * \sigma(x)$. Гиперпараметры α, β настраиваются в зависимости от типа данных, для данного датасета подобраны $\alpha = 5, \beta = 12$

Для полученной оценки лагранжиана можно сэмплировать значения функции в любой точке $(\mathbf{q}, \dot{\mathbf{q}})$. Соответственно сэмплируется норма, на основе которой строится метрика, а на основе этой метрики уже классифицируются траектории базовыми алгоритмами.

Поэтому для начала получается набор точек, в которых получают значения лагранжиана. Этот набор получается до классификации, и он един для всех траекторий. Соответственно он является гиперпараметром для классификатора. Получается набор $S = \{s_i\}$ из $N = 20000$ точек $(\mathbf{q}, \dot{\mathbf{q}})$ равномерно распределенных на кубе $-10 \leq x_i \leq 10$. Размер куба подбирается так, чтобы он покрывал точки, на которых обучаются лагранжианы. Число точек для сэмплирования должно быть подобрано так, чтобы оценка нормы менялась слабо по сравнению с ее оценкой.

Для каждого объекта получают значения $X_i = \{AL(s_i)\}$ набор его признаков. Норма в этом пространстве соответствует обычной евклидовой норме, а для классификации объектов в данном пространстве используются базовые алгоритмы классификации вето-ров. Классифицировать объекты предлагается по их признакам, т. е. сэмплированным значениям нормы.

$$A(L) = \frac{\partial L}{\partial \mathbf{q}} - \frac{\partial}{\partial \mathbf{q}} \frac{\partial L}{\partial \dot{\mathbf{q}}} \dot{\mathbf{q}}$$

Для регуляризации и уменьшения размерности применим метод главных векторов для преобразования признаков. После этого для исследования набора представим объекты по первым трем главным признакам на 2D и 3D-графиках.

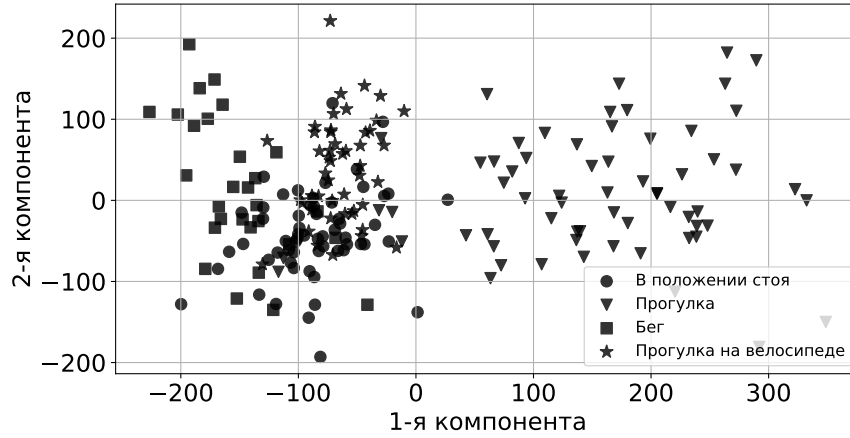


Рис. 5 Проекция оценок лагранжианов \hat{L} на 2D-поверхность

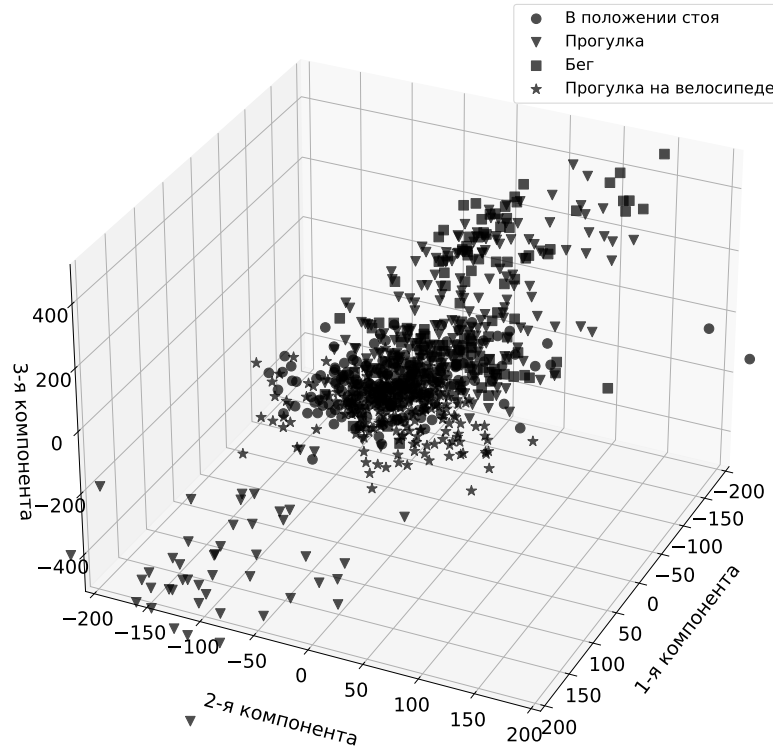


Рис. 6 Проекция оценок лагранжианов \hat{L} на 3D-поверхность

Из графиков 5, 6 следует, что данные разделяются на различные кластеры, распределенные по нормальному закону, и предполагается, что они разделяются метрическими методами с гауссовским ядром. Из графиков следует то, что в представленное пространство лагранжианов классы отделяются линейными поверхностями и следует использовать линейные модели для дальнейшей классификации.

Исследование моделей.

Для исследования был выбран случайный лес, т.к. он хорошо работает с любыми сложными наборами данных, линейная регрессия и гауссовский процесс, как достаточно простые классификаторы которые быстро учатся и соответственно полезно знать работают ли они на наборе данных, а также SVC с гауссовским ядром, т.к. по визуальному анализу предполагается, что данные распределены по нормальному закону.

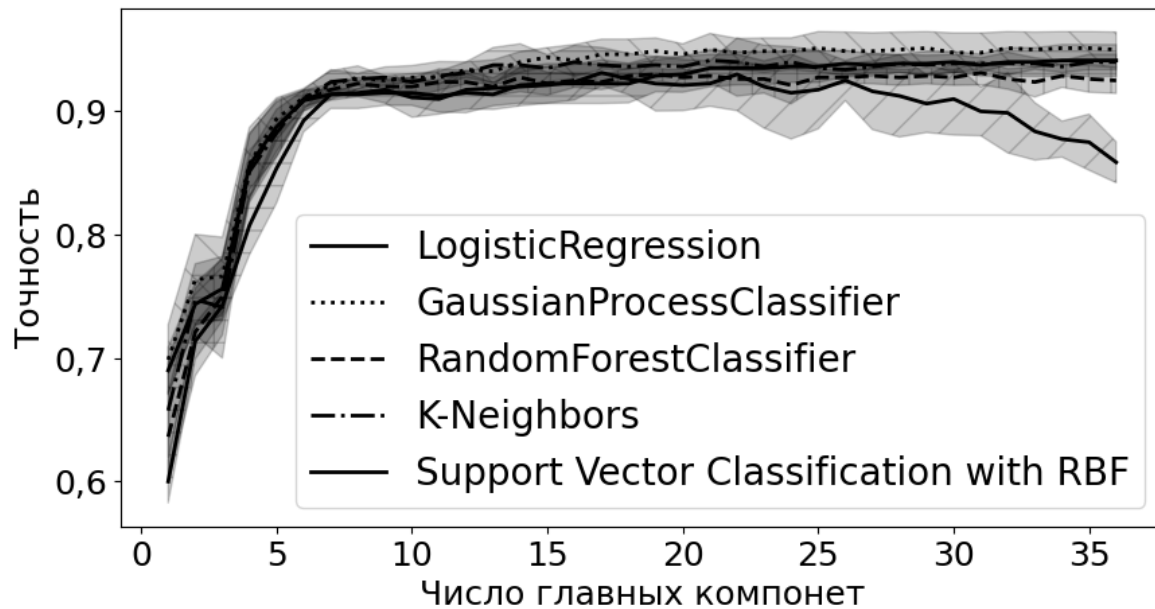


Рис. 7 Точность классификации выбранных методов в зависимости от числа главных компонент

Классификатор	Метрика		
	Accuracy	Balanced-accuracy	F1 Macro
Логистическая регрессия	0,927 ± 0,014	0,924 ± 0,13	0,927 ± 0,14
Гауссовский процесс	0,946 ± 0,010	0,941 ± 0,011	0,946 ± 0,010
Случайный лес	0,932 ± 0,007	0,928 ± 0,008	0,933 ± 0,008
К-ближайших соседей	0,939 ± 0,009	0,935 ± 0,010	0,940 ± 0,008
SVC с гауссовским ядром	0,933 ± 0,012	0,927 ± 0,013	0,933 ± 0,011

Таблица 1 Результат работ классификаторов на предложенной векторизации данных

Стабильнее всего обучается SVC, на нем отсутствует переобучение, и он показывает результаты сравнимые с остальными классификаторами, гауссовская переобучается, по-

этому абсолютно не подходит под модель. Гауссовский процесс и логистическая регрессия на пространстве больших размерностей. remark Случайный лес работает на уровне SVC

Сравнение с другими классификаторами

Также были проведены сравнения с существующими методами классификации временных рядов. Метод случайного леса для временных рядов `TimeSeriesForestClassifier`, представленный в библиотеке `Sktime` показывает на том же наборе данных точность в 78%, что намного хуже LNN-классификатора

8 Заключение

Представление траектории в виде лагранжианов физических систем, которым они соответствуют, предложенное в ранних работах, является мощным инструментом для классификации траекторий. Данный метод имеет множество преимуществ по сравнению с методами классификации, не использующие информацию о физике системы. Метод является многомерным уже по своей природе, нам не нужно вводить никаких предположений о связи между различными компонентами траекторий. Он устойчив к изменениям данных, сохраняющих физику системы. Предложенная векторизация данных сохраняет физические предположения о близости данных. В том числе метод дает удобный способ для проецирования данных на плоскость для дальнейшей визуализации. Визуальный анализ графиков показывает, что представленное пространство лагранжианов соответствует пространству признаков, которые задают класс траекторий, т.е. они определяют вид движения. Аналогичные результаты показывает обучение линейных моделей на данном векторном пространстве, из графиков следует, что с точностью 0.95 линейные модели разделяют классы, представленные в наборе данных. В то же время предложенная векторизация затратна по времени, а также требуется настройка параметров регуляризации для достижения необходимой точности аппроксимации на конкретных наборах траекторий

Литература

- [1] *Karniadakis, G.E., Kevrekidis, I.G., Lu, L. et al.* Physics-informed machine learning // *Nat Rev Phys* 3, 422–440 (2021). doi: <http://dx.doi.org/10.1038/s42254-021-00314-5>
- [2] *Ravinder Bhattoo et al.* Learning the dynamics of particle-based systems with Lagrangian graph neural networks // *Mach. Learn.: Sci. Technol.* 4 015003 (2023). doi: <http://dx.doi.org/10.1088/2632-2153/acb03e>
- [3] *Henning Wessels, Christian Weißenfels, Peter Wriggers* The neural particle method – An updated Lagrangian physics informed neural network for computational fluid dynamics // *Computer Methods in Applied Mechanics and Engineering*, Vol. 368 (2020), 113127. doi: <http://dx.doi.org/10.1016/j.cma.2020.113127>
- [4] *Nate Gruver, Marc Finzi, Samuel Stanton, Andrew Gordon Wilson* Deconstructing the Inductive Biases of Hamiltonian Neural Networks // *ICLR* (2022), doi: <http://dx.doi.org/10.48550/arXiv.2202.04836>
- [5] *Ruiz AP, Flynn M, Large J, Middlehurst M, Bagnall A.* The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. // *Data Min Knowl Discov.* (2021), doi: <http://dx.doi.org/10.1007/s10618-020-00727-3>
- [6] *Multivariate Time-Series Classification of Critical Events from Industrial Drying Hopper Operations: A Deep Learning Approach*
- [7] *S. Seto, W. Zhang and Y. Zhou* Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition // *IEEE Symposium Series on Computational Intelligence* pp. 1399-1406 (2015), doi: <http://dx.doi.org/10.1109/SSCI.2015.199>
- [8] *M. Cranmer, S. Greydanus, S. Hoyer et al.* Lagrangian neural networks // *ICLR 2020 Deep Differential Equations Workshop*. doi: <http://dx.doi.org/10.48550/arXiv.2003.04630>.
- [9] *Ismail Fawaz, H., Forestier, G., Weber, J. et al.* Deep learning for time series classification: a review // *Data Min Knowl, Disc* 33, 917–963 (2019) doi: <http://dx.doi.org/10.1007/s10618-019-00619-1>
- [10] *Dempster A, Petitjean F, Webb G* ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Min Knowl Disc* 34 1454–1495 (2020) doi: <http://dx.doi.org/10.1007/s10618-020-00701-z>
- [11] *Dempster A, Petitjean F, Webb G* Canonical and noncanonical Hamiltonian operator inference *Computer Methods in Applied Mechanics and Engineering*, Vol. 416 (2023) doi: <http://dx.doi.org/10.1016/j.cma.2023.116334>

- [12] *Dempster A, Petitjean F, Webb G* Adaptable Hamiltonian neural networks Phys. Rev. Res. Vol. 3 (2021) doi: <http://dx.doi.org/10.1103/PhysRevResearch.3.023156>
- [13] *Zakwan Muhammad, Xu Liang, ferrari trecate Giancarlo*. Robust Classification Using Contractive Hamiltonian Neural ODEs IEEE Control Systems Letters. 7. 145-150. (2021) doi: <http://dx.doi.org/10.1109/LCSYS.2022.3186959>
- [14] Physics-Informed Neural Network for Volumetric Sound field Reconstruction of Speech Signals
- [15] *Samuel Greydanus, Misko Dzamba, Jason Yosinski* Hamiltonian Neural Networks
- [16] *Kris Tucker, Amit Kiran Rege, Conor Smith, Claire Monteleoni, Tameem Albash* Hamiltonian Learning using Machine Learning Models Trained with Continuous Measurements