ПОЛИТЕХ
Институт кибербезопасности
и защиты информации

Санкт-Петербургский политехнический университет Петра Великого
Институт Кибербезопасности и Защиты Информации

ВЫЯВЛЕНИЕ АНОМАЛИЙ В РАБОТЕ КИБЕРФИЗИЧЕСКИХ СИСТЕМ НА ОСНОВЕ МЕХАНИЗМА ГИПЕРКУБА

А. Д. Фатин, Е. Ю. Павленко

Исследование выполнено в рамках стипендии Президента РФ молодым ученым и аспирантам СП-1689.2019.5

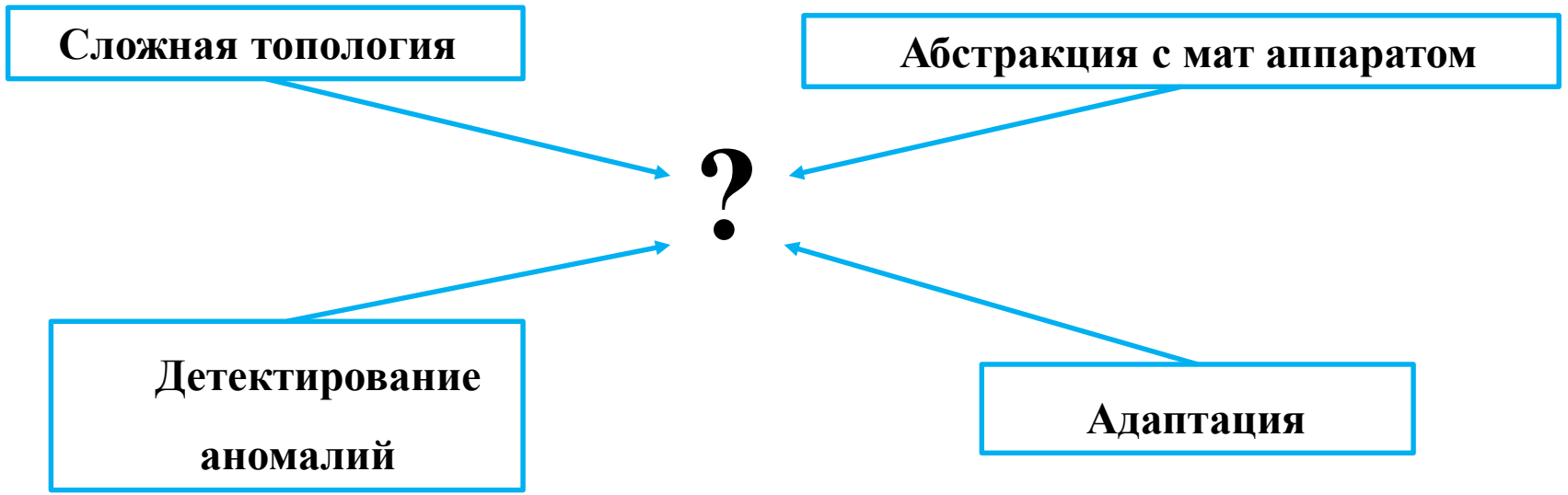
Постановка задачи

Киберфизическая система — информационно-технологическая **концепция**, подразумевающая интеграцию вычислительных ресурсов в физические сущности любого вида.

В КФС **вычислительная компонента распределена** по всей физической системе, которая является её носителем, **и синергетически увязана с** её **составляющими** элементами.

Математическая модель — математическое **представление** некой сущности, один из вариантов модели как системы, исследование которой позволяет **получать информацию** о некоторой другой системе.

Математическая модель предназначена **предсказать поведение** реального объекта, но всегда **представляет** собой ту или иную **степень** его **идеализации**.



Рассматриваемая область

Алгоритм действий

Нижний логический уровень КФС

↓

Данные

↓

Многомерные временные ряды

↓

Некое преобразование

↓

Предсказание и/или анализ состояния

Модель

Многомерный временной ряд – **вектор** векторов произольной длины, **зависимый от времени**.

$X = \{X^{(1)}, X^{(2)}, \dots, X^{(m)}\}$, где каждое значение в момент времени t_i представлено вектором: $X = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$.

Матрица? Не всегда.

Нормировка:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Основные черты математических моделей КФС

Задача\Модель	Временные ряды	Алгоритм Калмана	ДВП	Фракталы	Графы
Вариативность	+	+	+	-	+
Краткосрочные атаки	+	+	+	+	+
Долгосрочные атаки	-	+	-	+	+
Ручная настройка	+	-	-	+	+/-
Агрегация данных	-	+	+	+	+/-
Производительность	+/-	+	+	+	-
Учет нелинейных процессов	-	+	-	+	+/-
Учет топологии системы	-	-	-	+	+
Унификация задачи	+	-	-	-	-

Машинное обучение

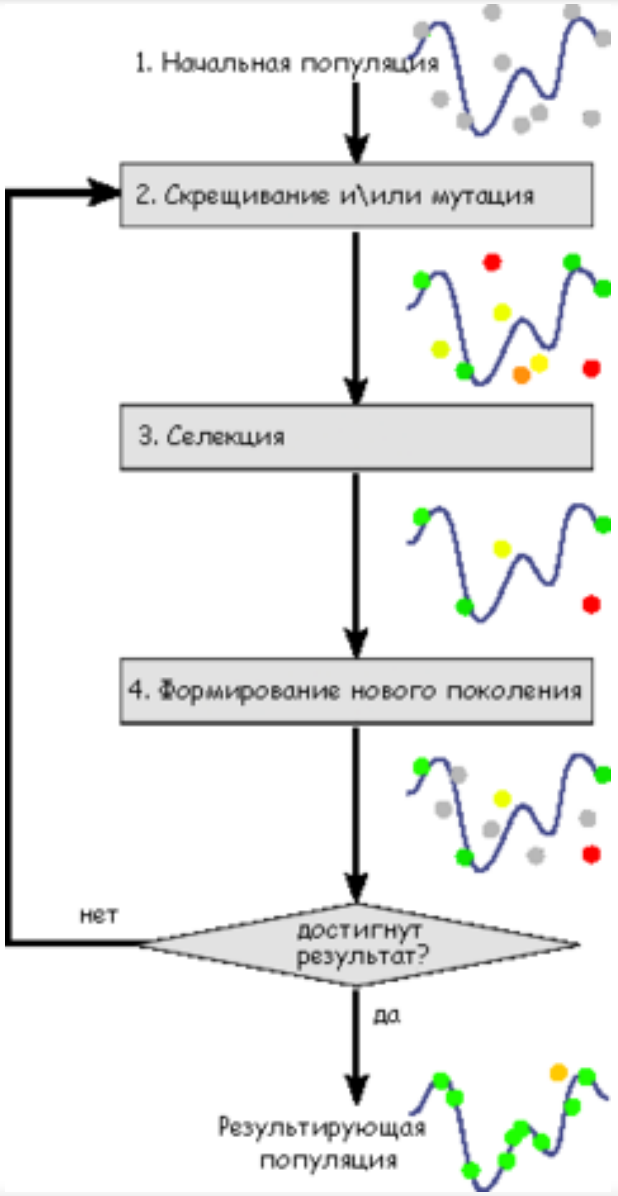
Генетические алгоритмы

Эвристический алгоритм поиска, используемый для решения задач **оптимизации** и **моделирования** путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, **аналогичных естественному отбору** в природе.

Вырождение решения в локальных экстремумах;

Оптимизация, но **не** предсказание;

Малое количество решаемых задач в сфере **предсказания**.



Нейронные сети

Некая математическая **модель**, а также её программное или аппаратное воплощение, построенная **по принципу** организации и **функционирования биологических нейронных сетей** — сетей нервных клеток живого организма.

Идеально подходит для **предсказания** состояния;

Изобилие всевозможных **видов** и конфигураций;

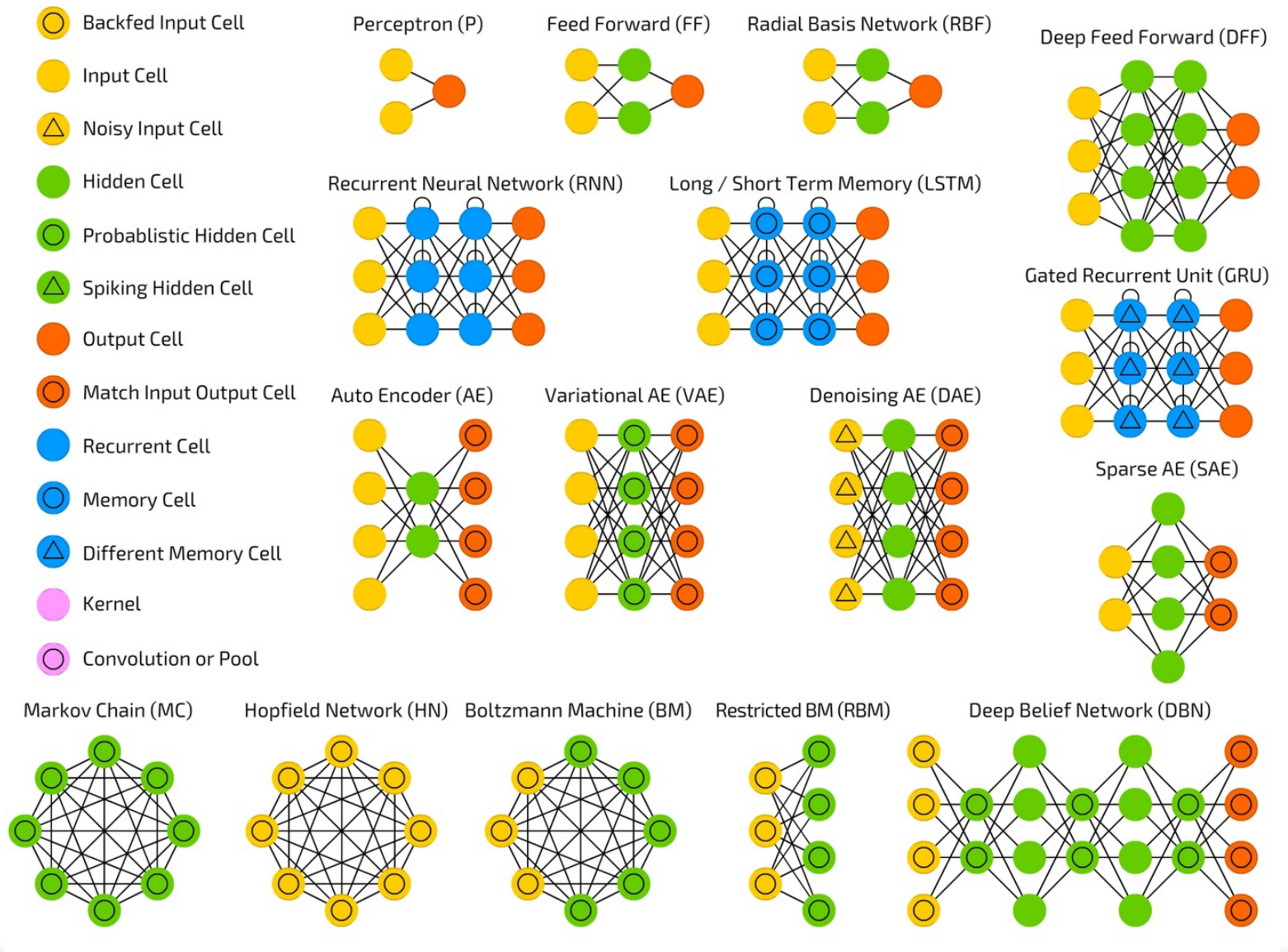
Огромный существующий научный **базис**.

Сложность проектирования;

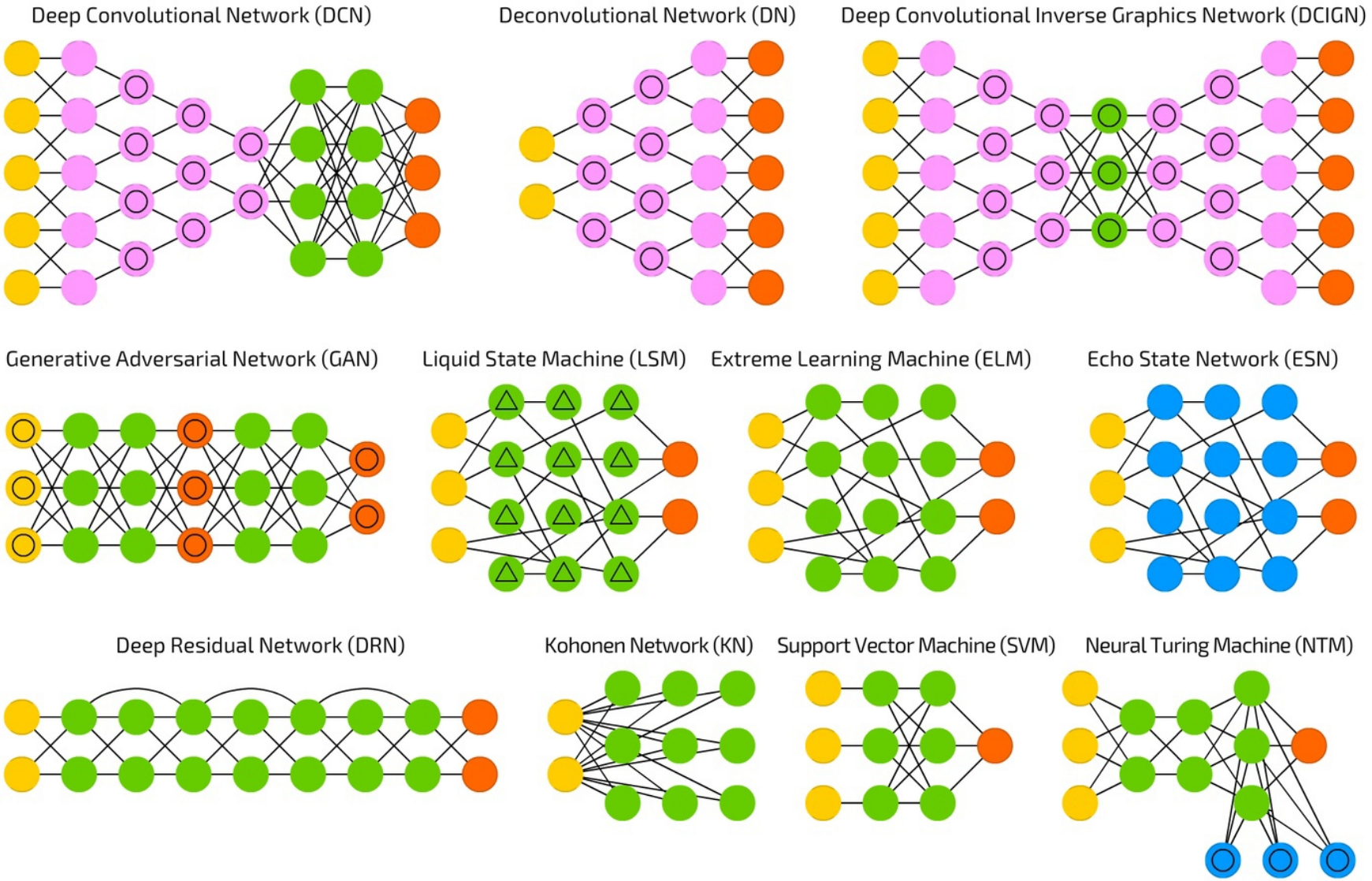
Вероятность **расхождения**;

Проблема **выбора**.

Проблема выбора топологии нейронной сети

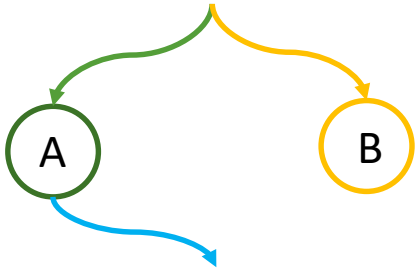


Проблема выбора топологии нейронной сети

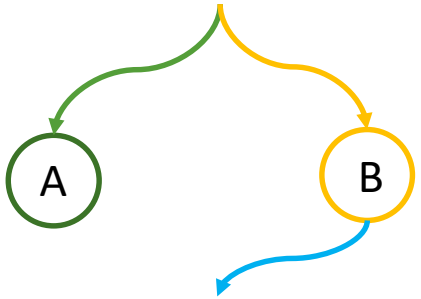


Нейрогенетические алгоритмы: общий принцип работы

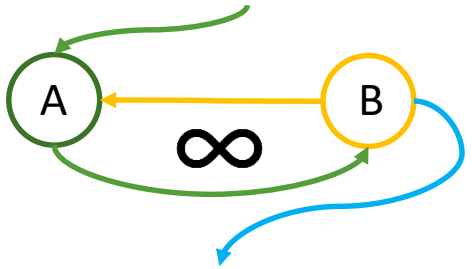
Генетические алгоритмы



Нейронные сети



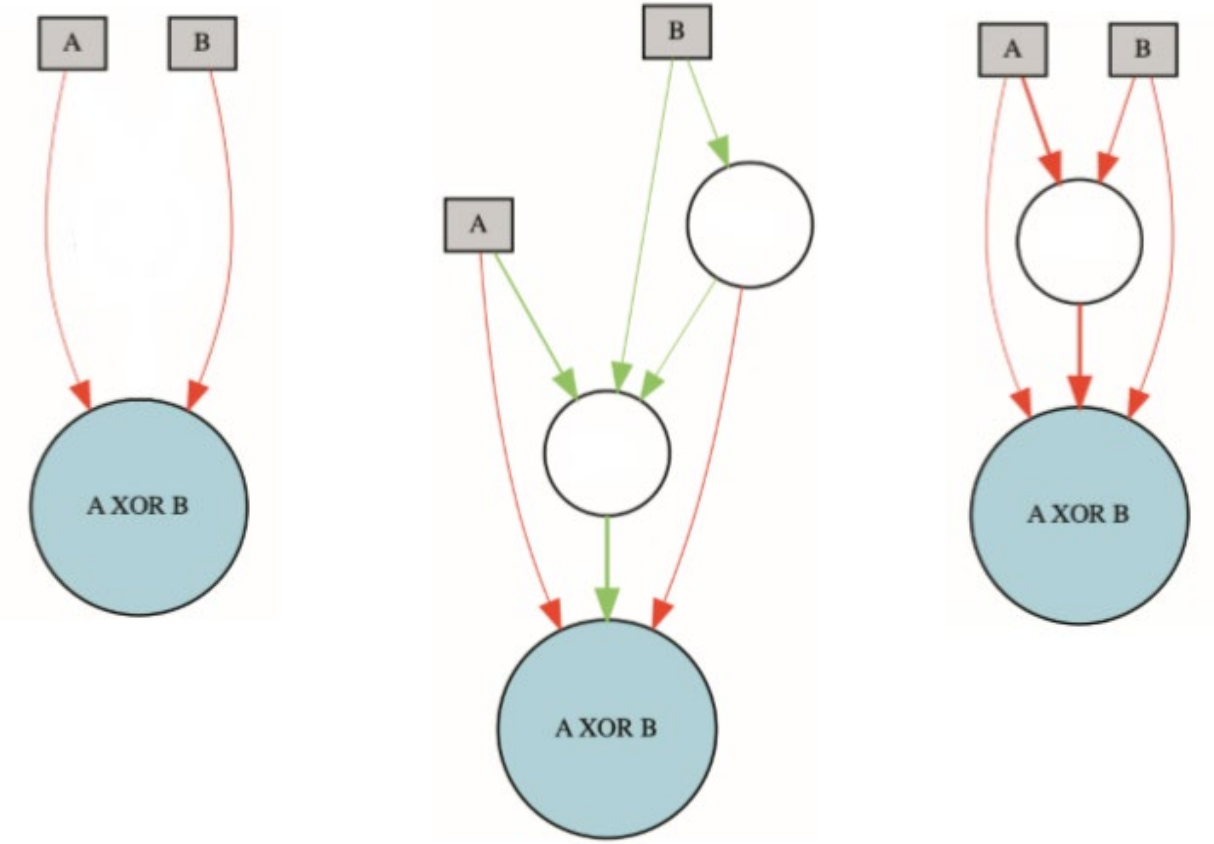
Нейрогенетические алгоритмы



Быстро, но **совсем не** точно	Средне, но **не** совсем точно	Долго, но максимально **точно**

Мутации как способ решения: простейший пример

Три этапа мутации решения задачи XOR'а



Посредством **мутации** топологии удалось найти **решение задачи с** изначально **некорректно подобранной конфигурацией**

Выгода обоих подходов;

Упрощение реализации на практике.

Время сходимости задачи **возрастает**.

Типы используемых операторов мутации и кроссовера

Инверсия

Изменение порядка

Изменение значения

Изменение экспрессии

Одноточечный кроссовер

Двухточечный кроссовер

Унифицированный (случайный) кроссовер

Генотип

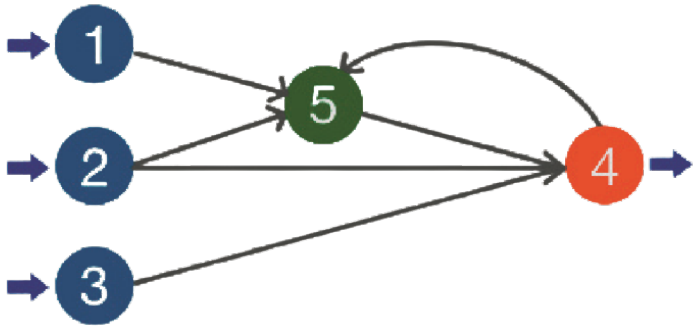
Гены узлов

Узел 1	Узел 2	Узел 3	Узел 4	Узел 5
Вход	Вход	Вход	Выход	Скрытый

Гены связей

Вход: 1	Вход: 2	Вход: 3	Вход: 1	Вход: 5	Вход: 2	Вход: 4
Выход: 4	Выход: 4	Выход: 4	Выход: 5	Выход: 4	Выход: 5	Выход: 5
Вес: 0,5	Вес: 0,7	Вес: 0,4	Вес: 0,6	Вес: 0,1	Вес: 0,3	Вес: 0,8
Выкл.	Вкл.	Вкл.	Вкл.	Вкл.	Вкл.	Вкл.

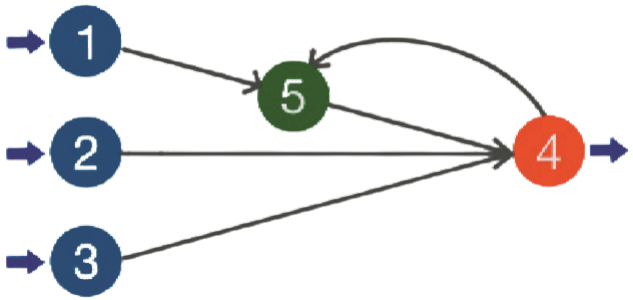
Фенотип



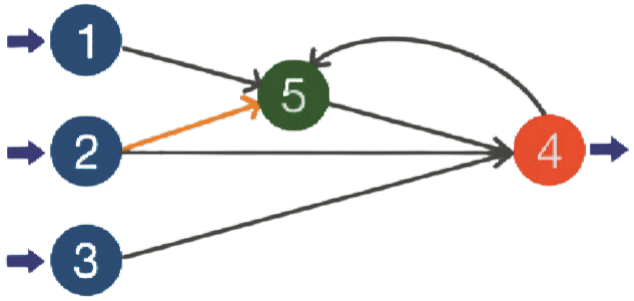
Конечные типы операторов мутаций

Мутации добавлением связей

1	2	3	4	5	6
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	4 -> 5

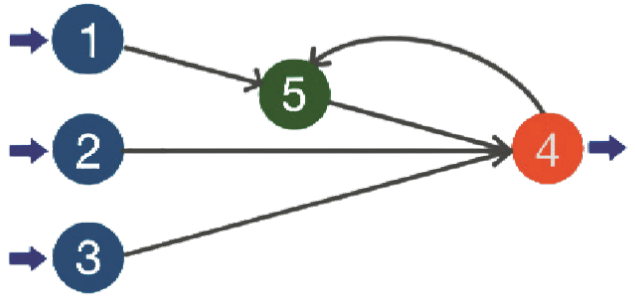


1	2	3	4	5	6	7
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	4 -> 5	2 -> 5

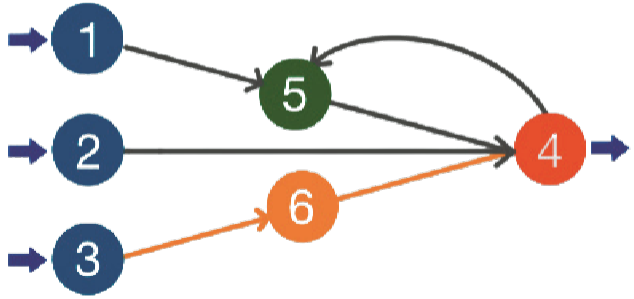


Мутации добавлением узла

1	2	3	4	5	6
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	4 -> 5



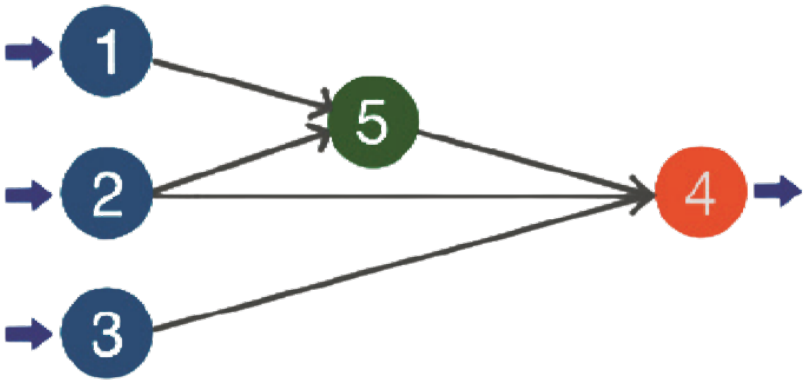
1	2	3	4	5	6	7	8
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	4 -> 5	3 -> 6	6 -> 4

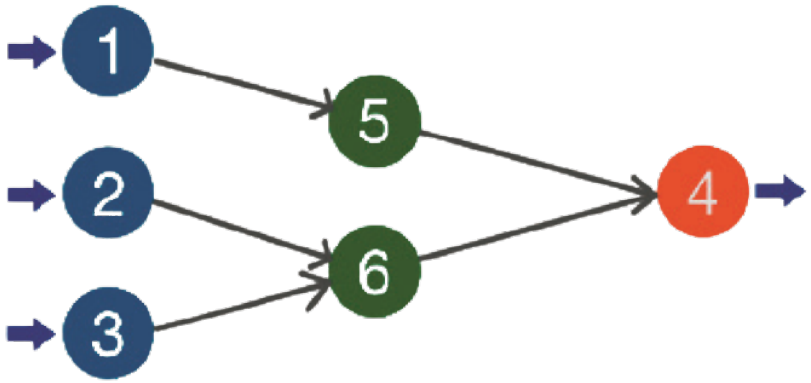


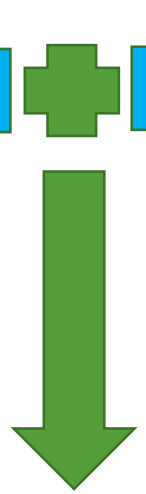
Конечный тип кроссовера

1	2	3	4	5	6	7	8
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4			2 -> 5

1	2	3	4	5	6	7	8	9
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	3 -> 6	6 -> 4		2 -> 6

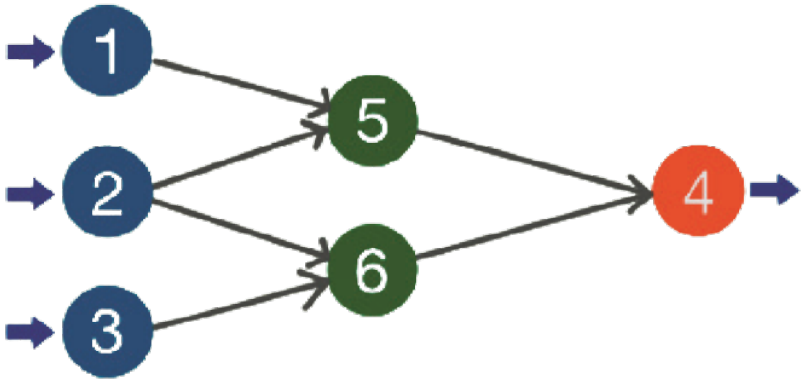






1	2	3	4	5	6	7	8
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4			2 -> 5

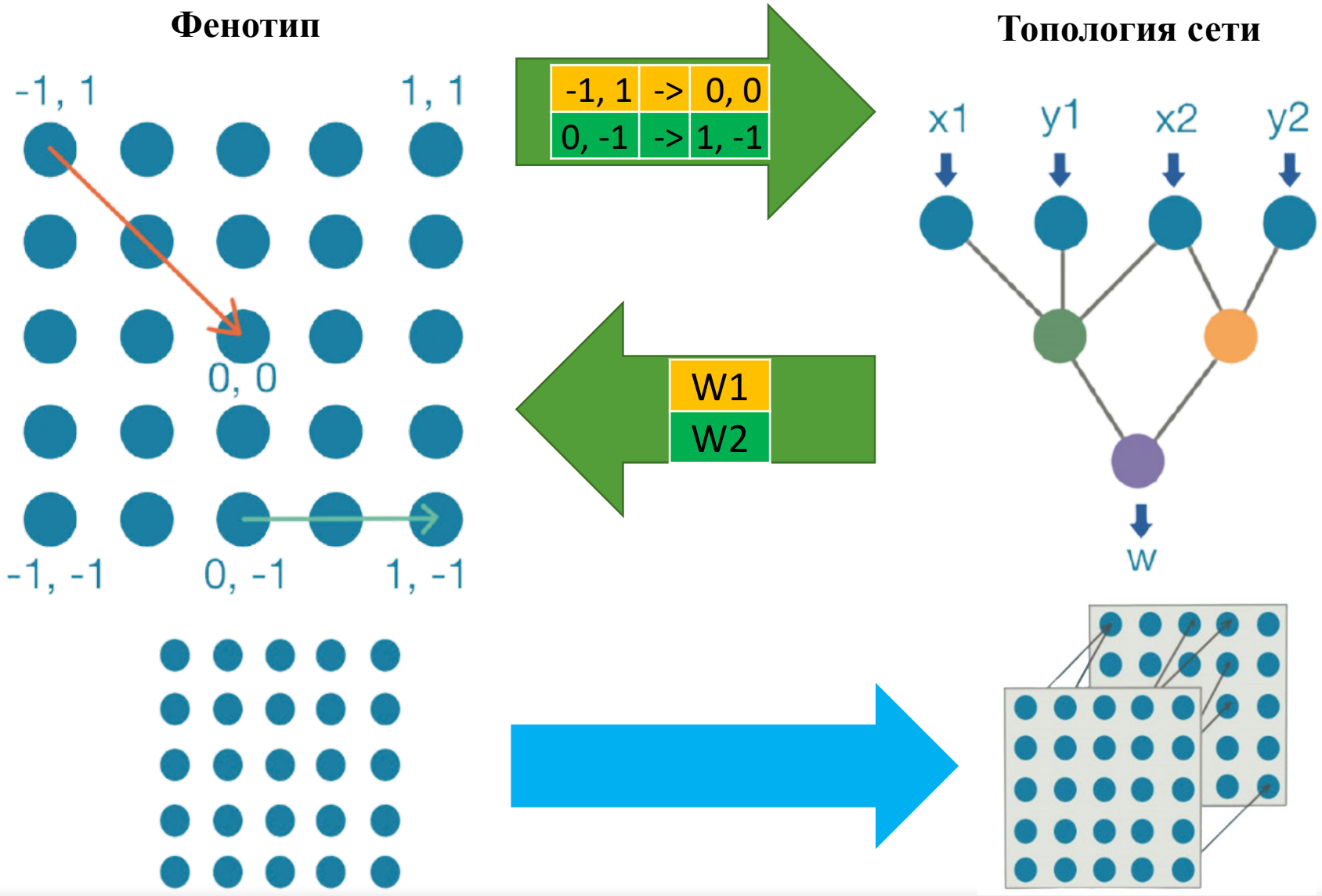
1	2	3	4	5	6	7	8	9
1 -> 4	2 -> 4	3 -> 4	1 -> 5	5 -> 4	3 -> 6	6 -> 4		2 -> 6



N-мерный гиперкуб

N-мерный гиперкуб:
w = f(x1, y1, x2, y2, …, g, …)

- **Различные** функции активации
- Слоистая структура
- **Сужение** областей **решения** за счёт **минимального размера** фенотипа
- **Линейность** отображения решения



Алгоритм *N*-мерного гиперкуба

begin

1. Выбор нужной конфигурации субстрата;
2. Инициализация популяции случайными весами связей;

repeat

for each *организма* из *популяции* **do**

3. Мутировать;
4. Выполнить операцию кроссовера

5. Оценить пригодность фенотипа целевой функцией;

until решение нейросети стагнирует

Исходные данные исследования

Система:
- TON_IOT DATASETS (https://ieee-dataport.org/documents/toniot-datasets)
- 7 устройств с 4 базисами
- 4 дискретных периода функционирования по 48 часов

Временной ряд:
- Классическое нормирование
- Δt шаг = 1 секунда
- 4 базиса каждого устройства
- Размерность ряда = 7x4 = 28

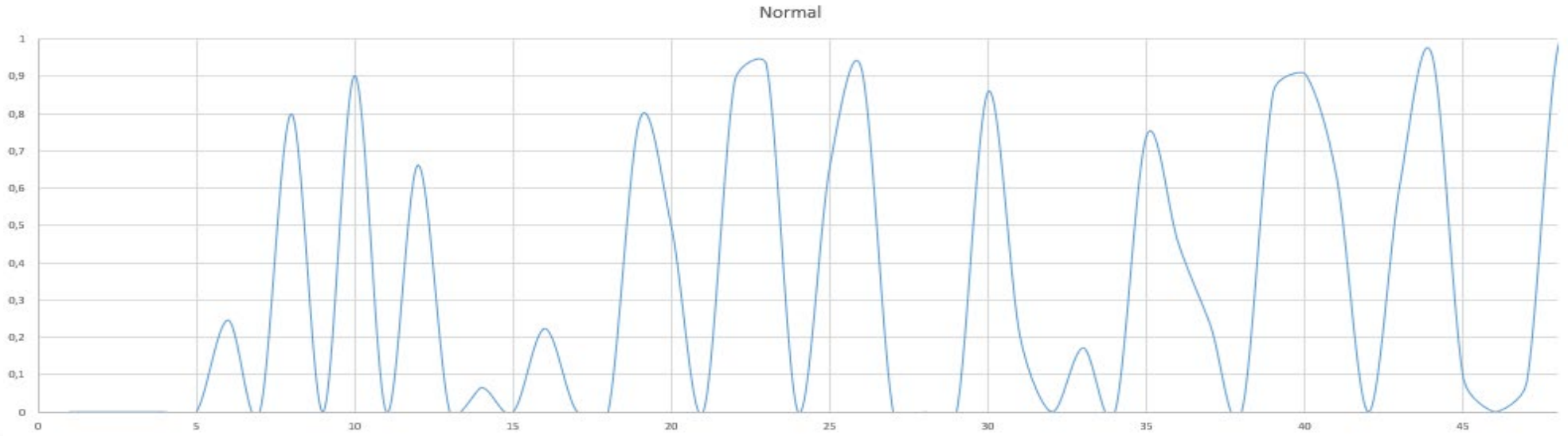
Типы рассмотренных атак:
- DoS
- DDoS
- Backdoor

Пороговая ошибка:
- Err = 0,4

Функция активации:
- Сигмоида

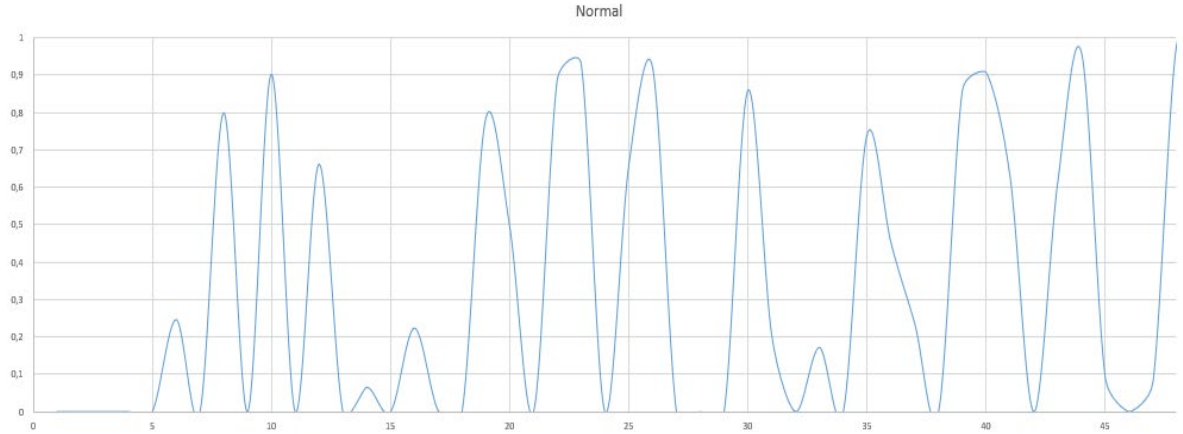
Предсказание:
- ***N***-мерный гиперкуб
- **12** слоев
- **744** поколения

Нормальное изменение данных:

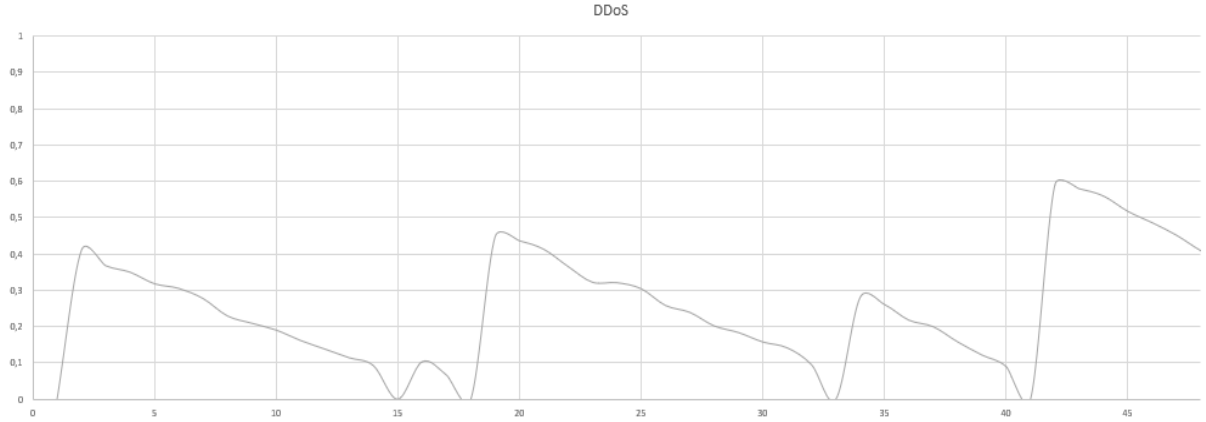


Графики изменения данных

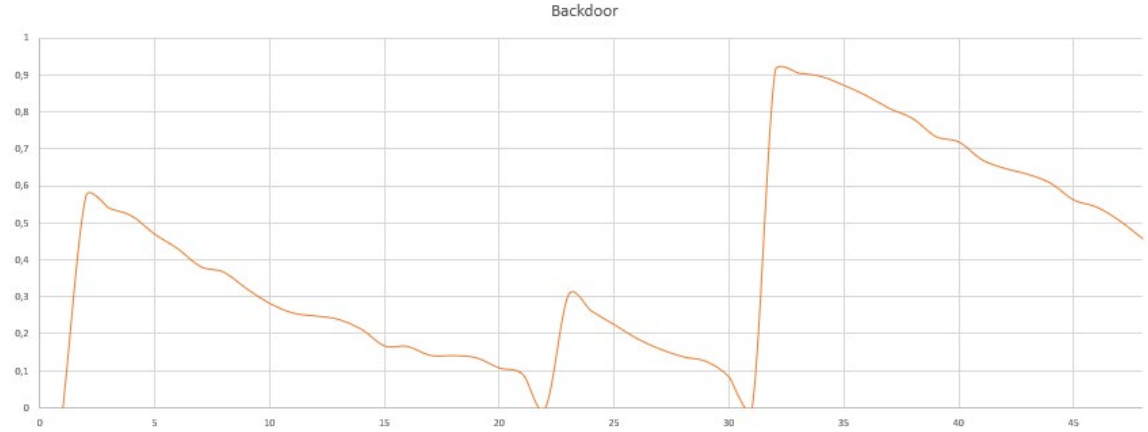
Нормальное изменение данных:



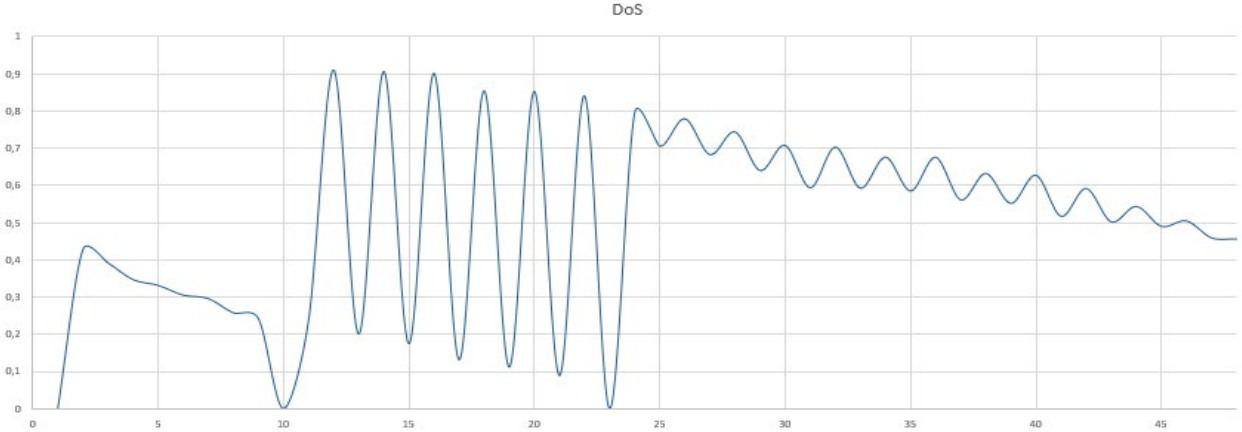
DDoS:



Backdoor:

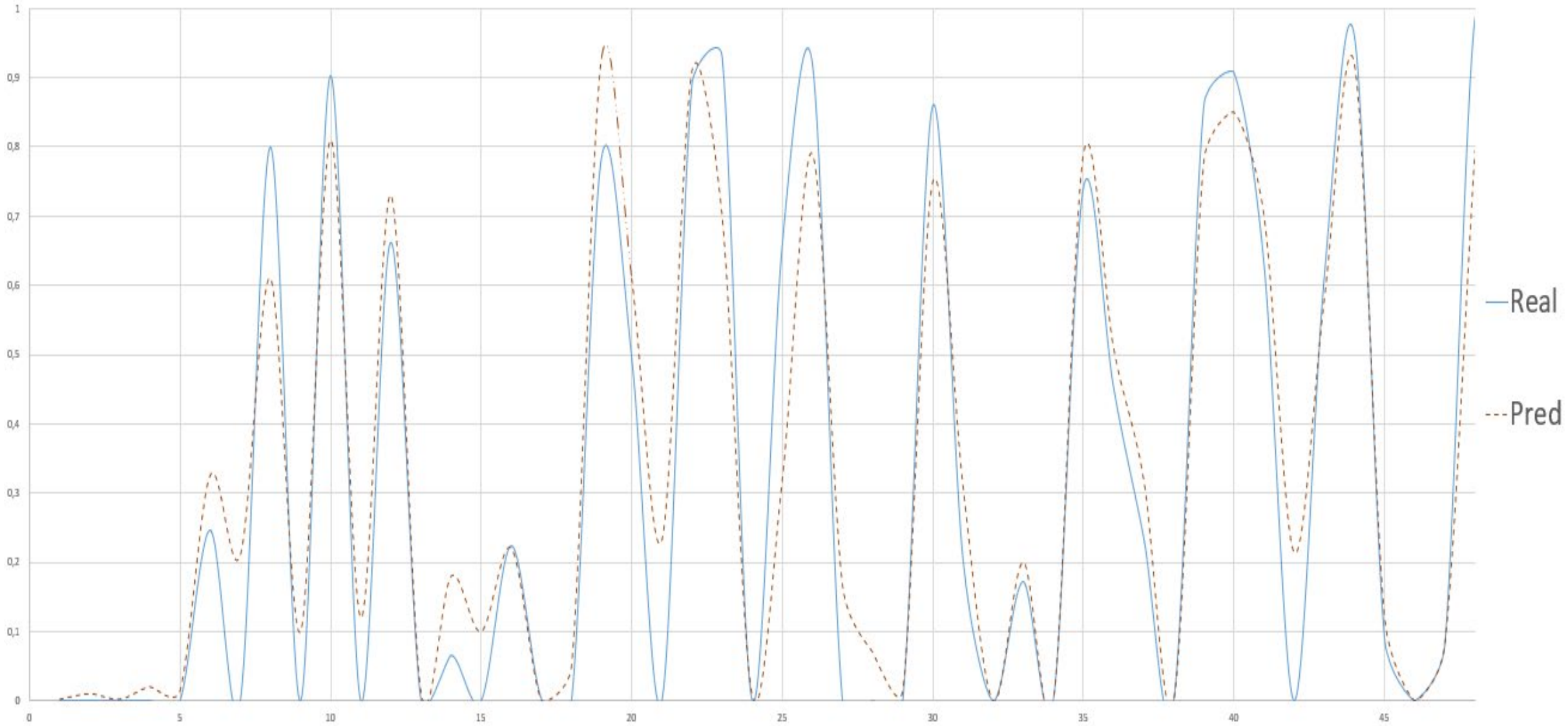


DoS:



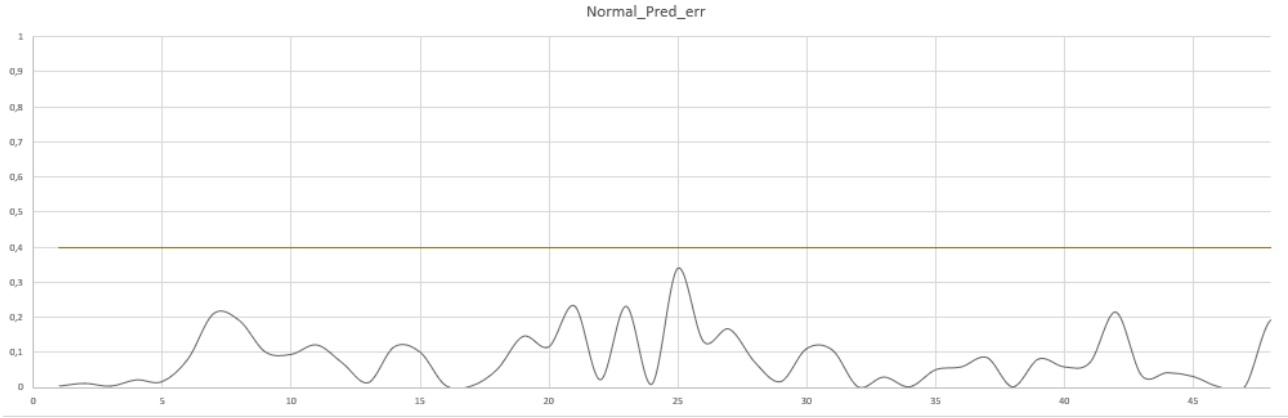
ПОЛИТЕХ

Предсказание состояния системы

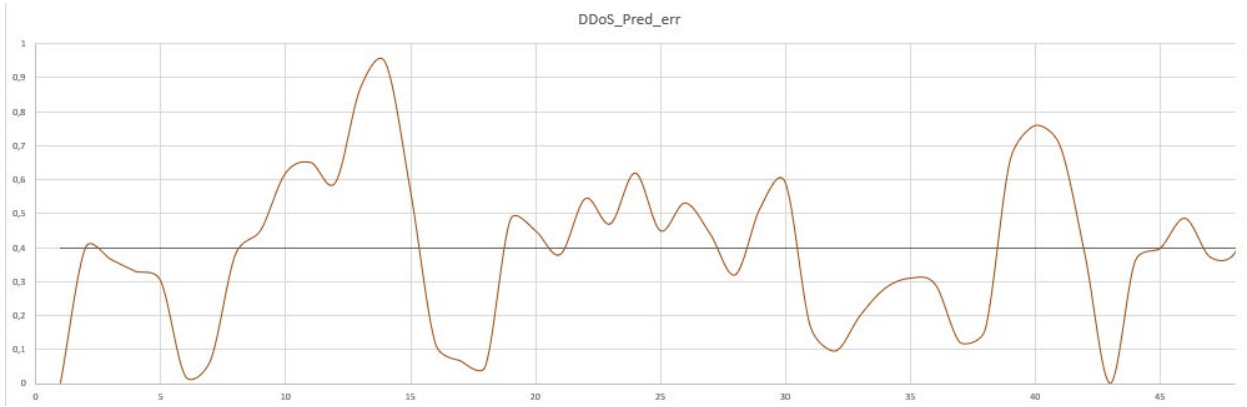


Ошибки предсказания системы для атак и для норм

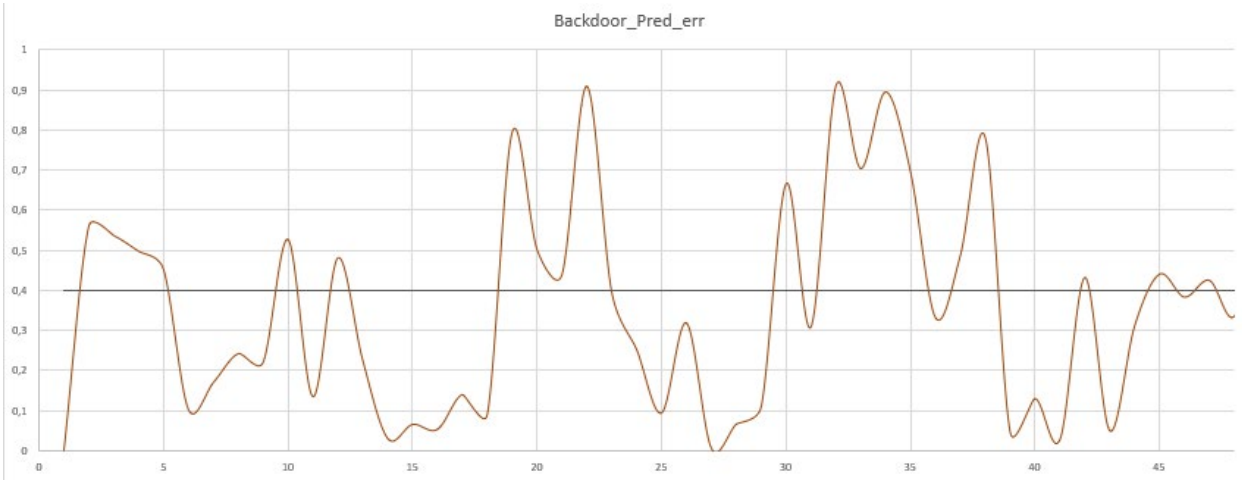
Нормальное изменение данных:



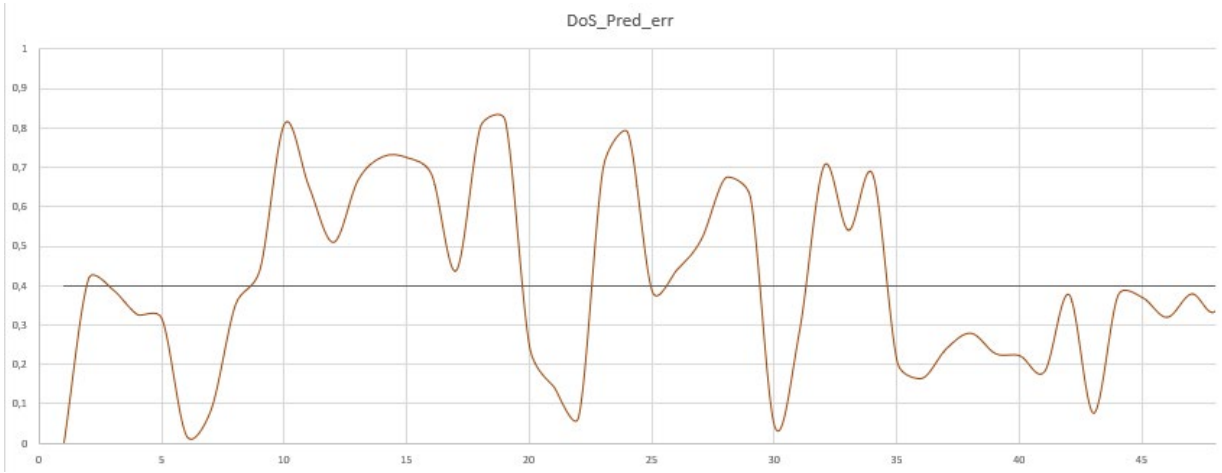
DDoS:



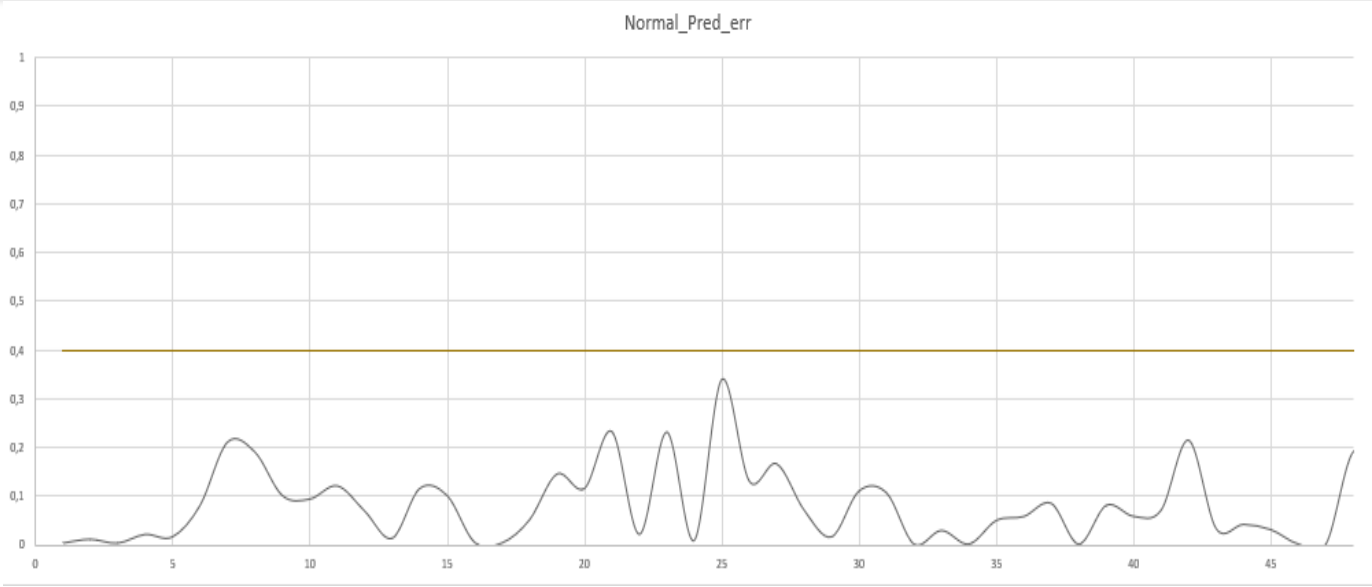
Backdoor:

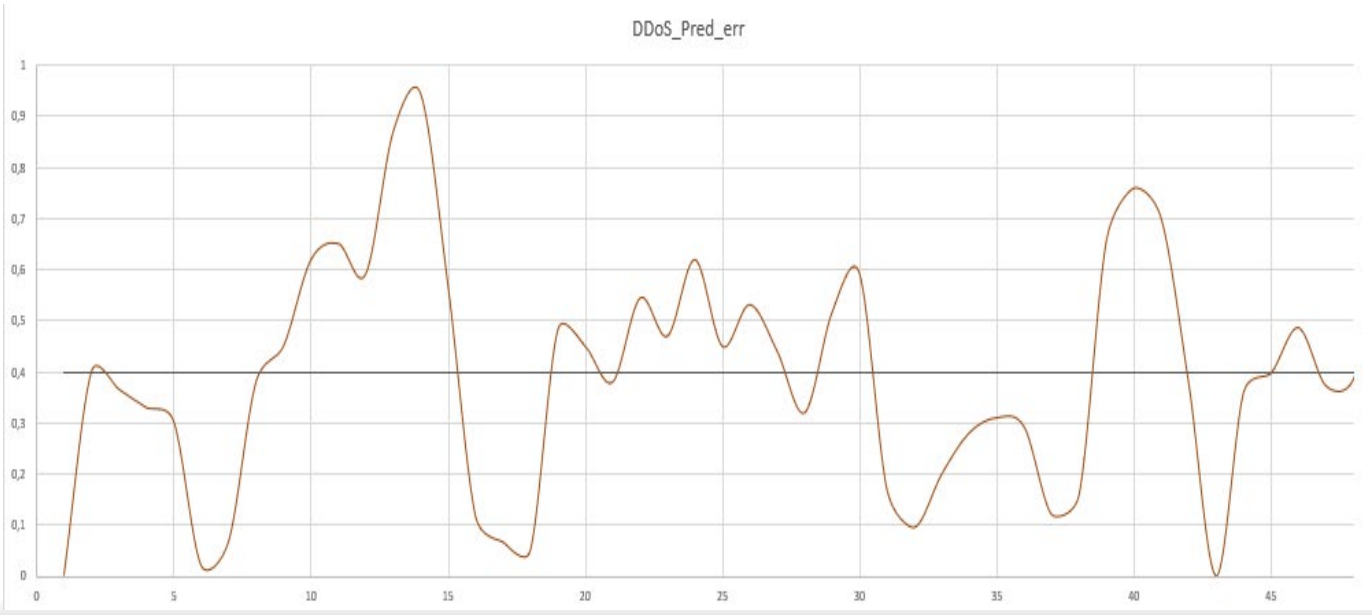


DoS:



Ошибки предсказания системы





При использовании **CNN-нейросети**:

- Точность > 84%
- Ошибка I рода < 0,13
- Ошибка II рода < 0,07

При использовании ***N***-мерного гиперкуба:

- Точность > 90%
- Ошибка I рода < 0,12
- Ошибка II рода < 0,01

Спасибо за внимание!