конференция
РусКрипто

Ежегодная международная научно-практическая конференция

«РусКрипто'2023»

Исследование эффективности применения нейросетевых алгоритмов для оценки минимальной энтропии последовательностей, вырабатываемых датчиками случайных чисел

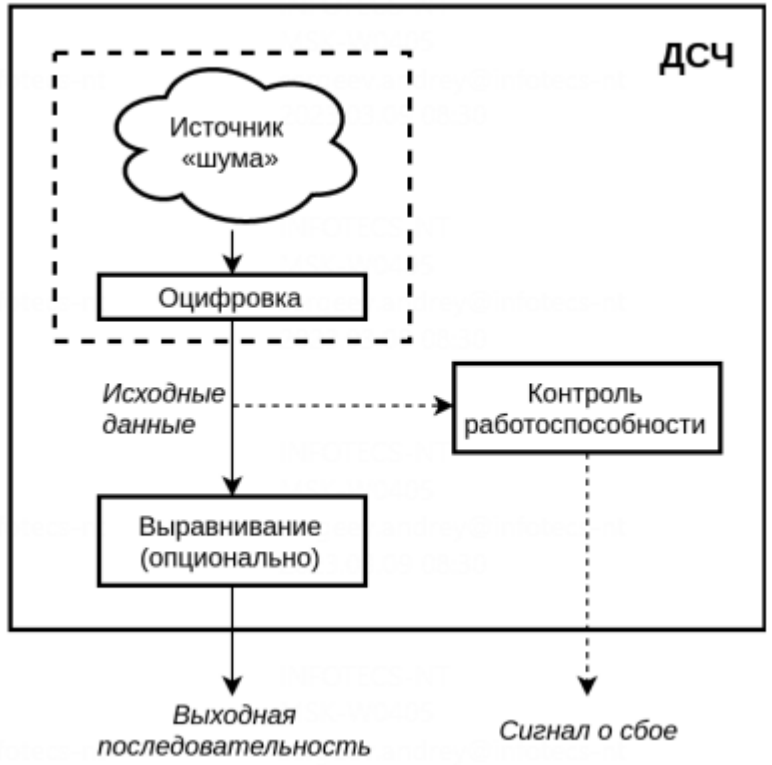
Сергеев Андрей,
Специалист, СФБ Лаб

Введение

- Задача оценки статистических характеристик датчиков случайных чисел
- Тесты NIST 800-22, Diehard, TestU01 хорошо подходят для ПДСЧ

Модель ФДСЧ, БиоДСЧ

Типичный ФДСЧ без усложнений «не пройдёт» тесты NIST и аналогичные



NIST 800-90B (2018 г.)
Требования к источнику «шума»

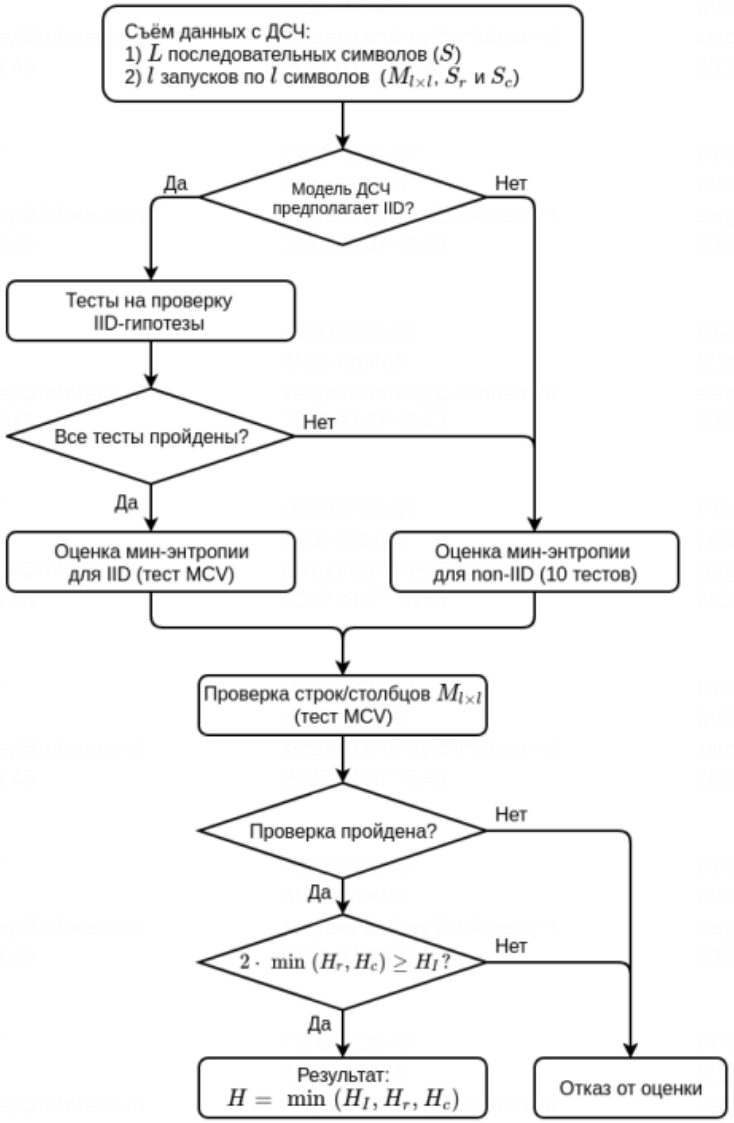
1. Стационарность выходной энтропии
2. Оценка мин-энтропии, обоснованная технической аргументацией и мат.моделью
3. Защита от активного влияния и пассивного наблюдения противником

От источника требуется не «идеальная случайность», а **высокая энтропийность**.

Схема работы NIST 800-90B

Две группы тестов:

- IID (independent and identically distributed) – независимые и одинаково распределенные случайные величины
- **Non-IID – сложные зависимости и нестационарные распределения**



Логика работы алгоритмов-предикторов

Тестируется последовательность из $N+M$ символов

Алгоритм-предиктор:

- получает на вход N символов
- пытается угадать N+1-й символ
- получает N+1 символов
- пытается угадать N+2-й символ
- …
- пытается угадать N+M-й символ

Число правильных угадываний – C

Доля правильных угадываний – $\mathrm{P} = \frac{C}{M}$

Оценка мин-энтропии «на один символ» $\mathrm{H} \approx -\log_2 P$

Оценка мин-энтропии

- Глобальный критерий – по средней вероятности угадывания
- Локальный критерий – по максимальной последовательности правильных угадываний
- Общая оценка – минимальная из глобальной и локальной оценки

Существующие алгоритмы-предикторы

NIST SP 800-90B:

- Multi Most Common in Window Prediction (MultiMCW) – самый частый символ
- Lag Prediction – проверка периода
- Multi Markov Chains (MultiMMC) Prediction – марковская цепь
- LZ78Y Prediction – на основе словарного подхода и соответствующего алгоритма сжатия

Цель исследовательской работы

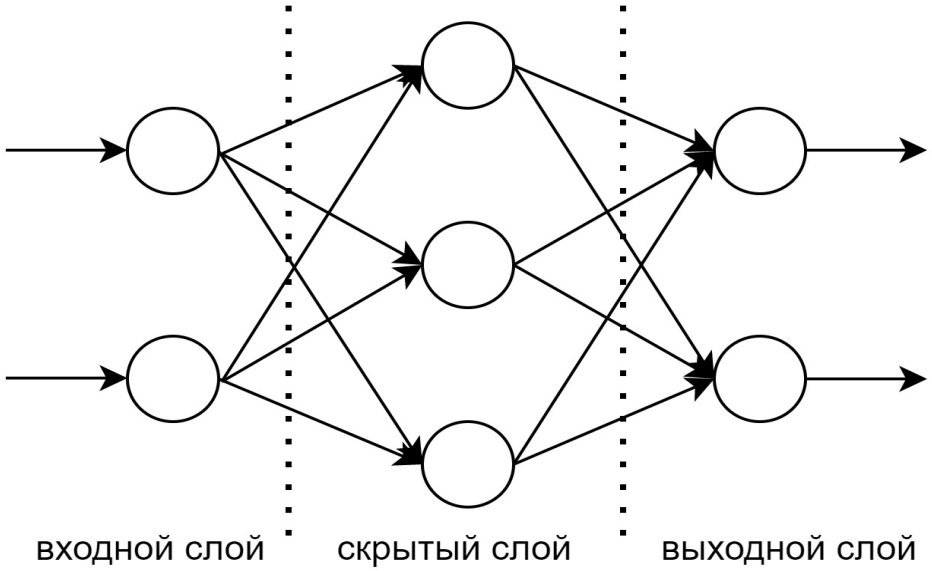
Проверить нейросетевые алгоритмы [1,2], предназначенные для предсказывания выходной последовательности, формируемой алгоритмическим датчиком с неизвестной структурой.

- [1] Na Lv, Tianyu Chen, Shuangyi Zhu, Jing Yang, Yuan Ma, Jiwu Jing, Jingqiang Lin, "High-Efficiency Min-Entropy Estimation Based on Neural Network for Random Number Generators", Security and Communication Networks, 2020.
- [2] Jing Yang, Shuangyi Zhu, Na Lv, Tianyu Chen, Yuan Ma, Jiwu Jing, Jingqiang Lin, "Erratum to “High-Efficiency Min-Entropy Estimation Based on Neural Network for Random Number Generators”", Security and Communication Networks, 2020.

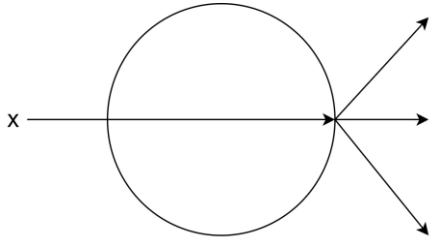
Архитектуры нейронных сетей

- Сети прямого распространения (FNN)
- Residual Network (ResNet)
- Рекуррентные нейронные сети (RNN)
- Сети встречного распространения (CPN)

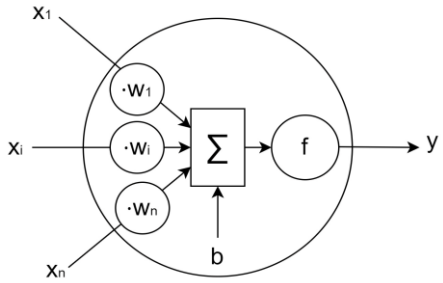
Сети прямого распространения FNN



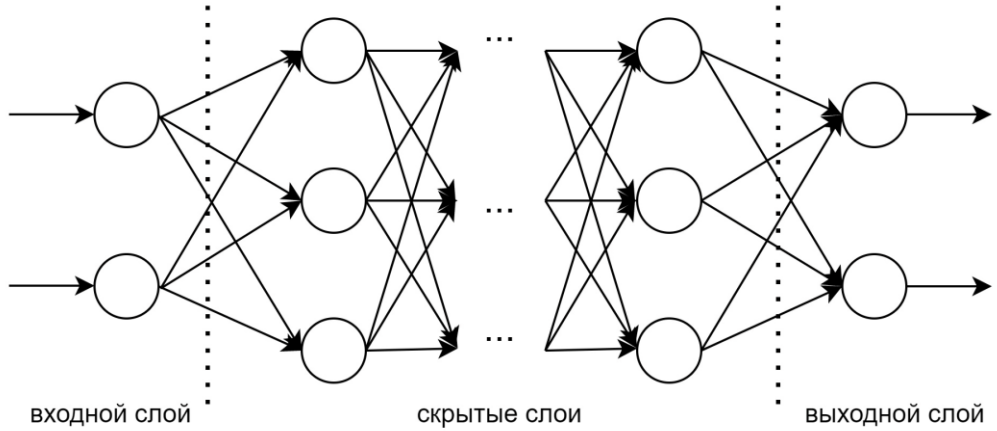
Строение сети типа FNN



Нейрон входного слоя

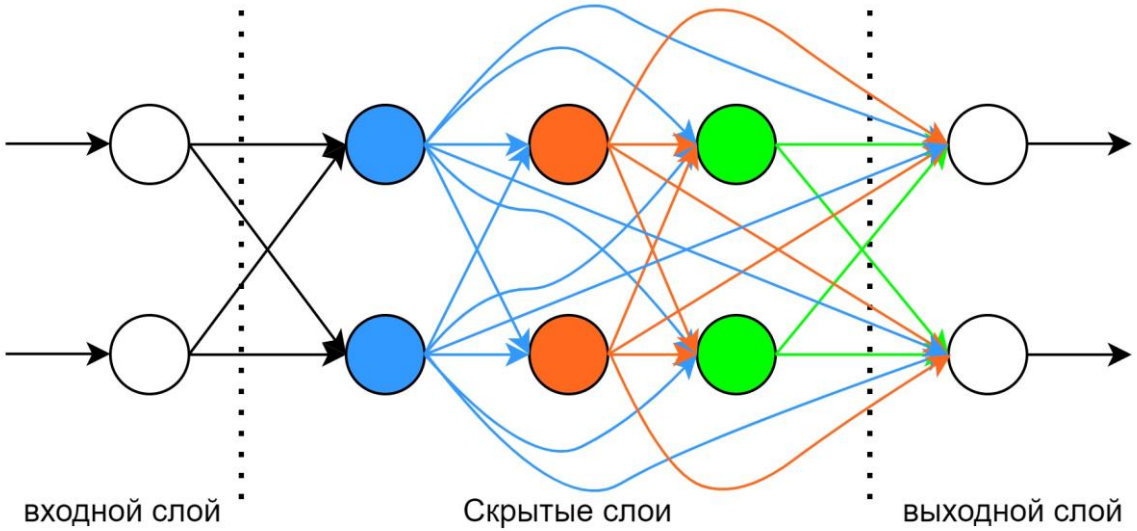


Нейрон скрытого и выходного слоя



Строение сети типа DNN

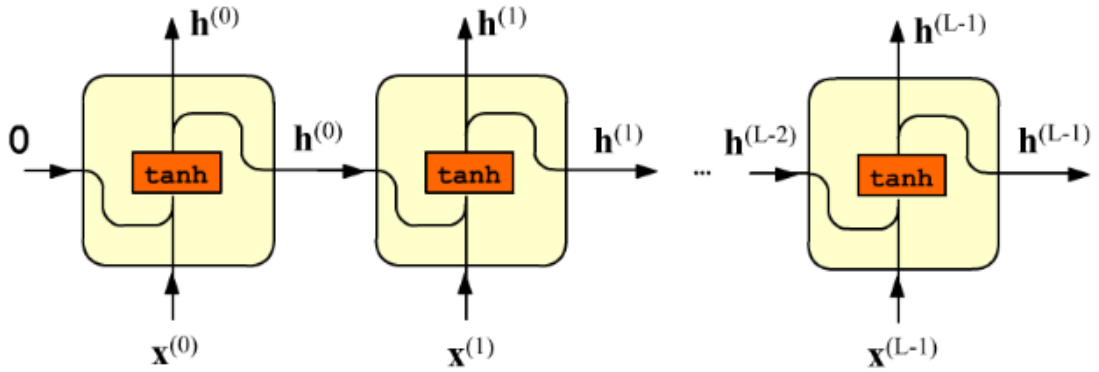
Residual Network ResNet



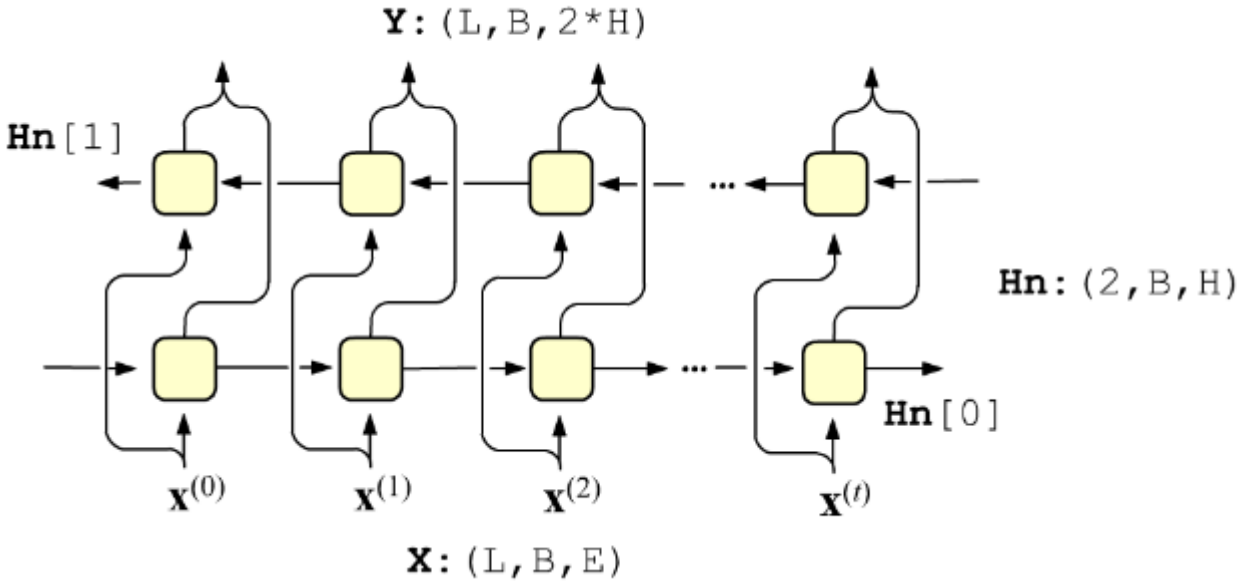
Нейронная сеть типа ResNet

Рекуррентные нейронные сети RNN

$$h^{(t)} = tanh(x^{(t)} \cdot W + h^{(t-1)} \cdot H + b)$$

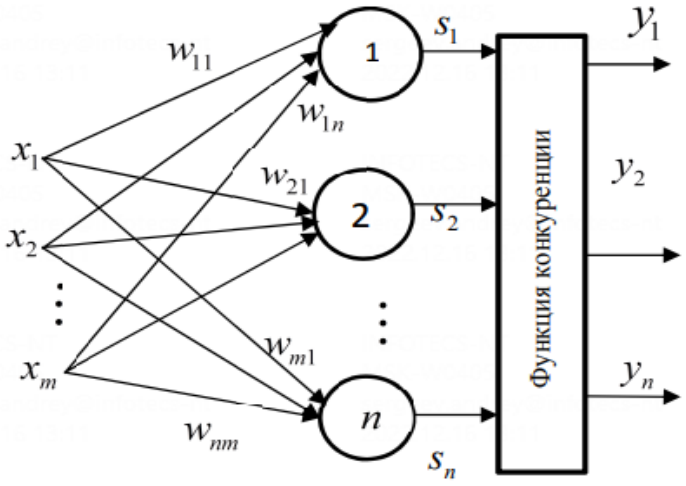


Простая RNN



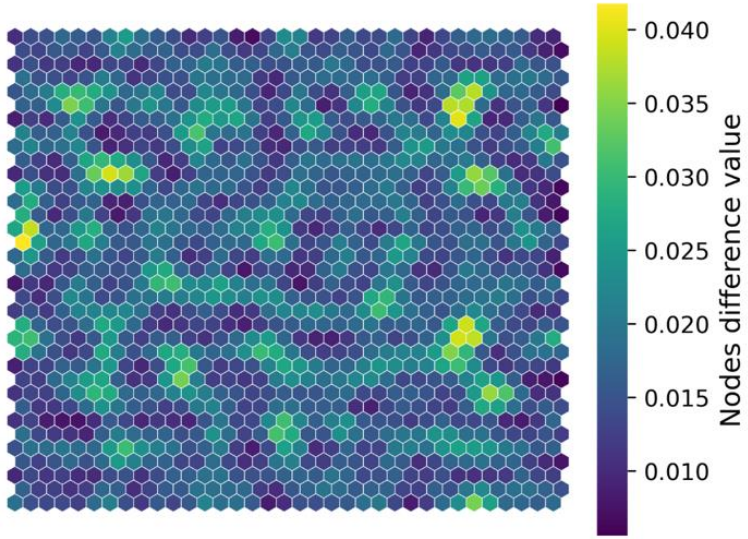
Двунаправленная RNN

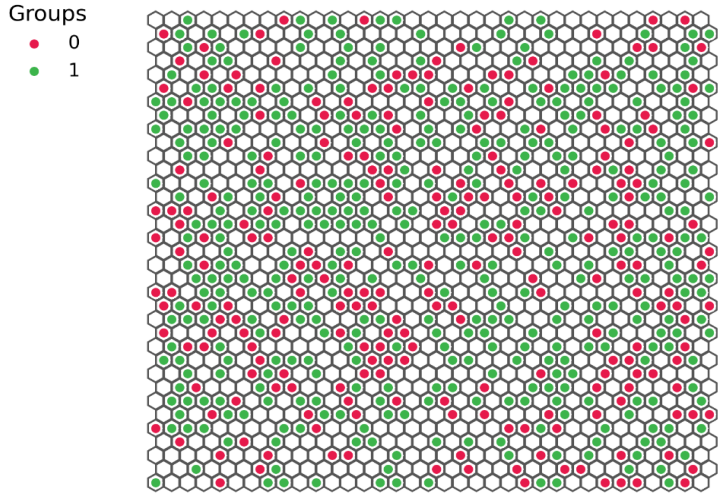
Сети встречного распространения CPN



Сеть Кохонена

Слой Гроссберга





Подготовка входных данных перед подачей в нейронную сеть

[0,1,2]

[(1,-1,-1), (-1,1,-1), (-1,-1,1)]

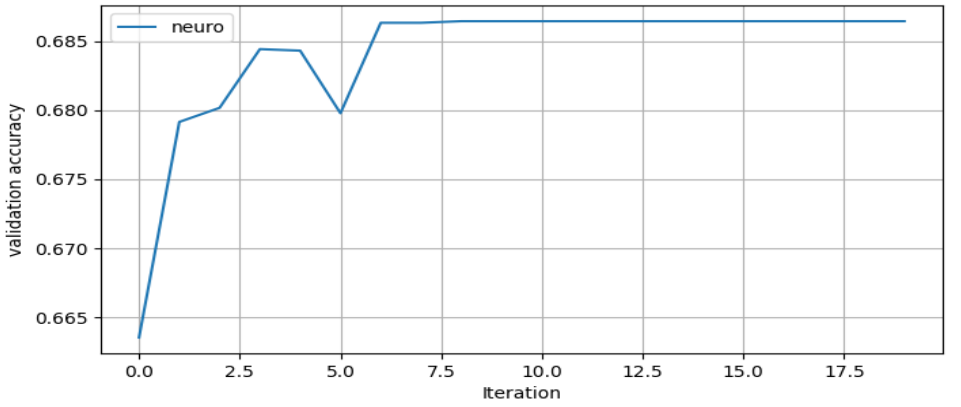
Тестовая программная реализация разработанных предикторов. Марковский источник глубины 10

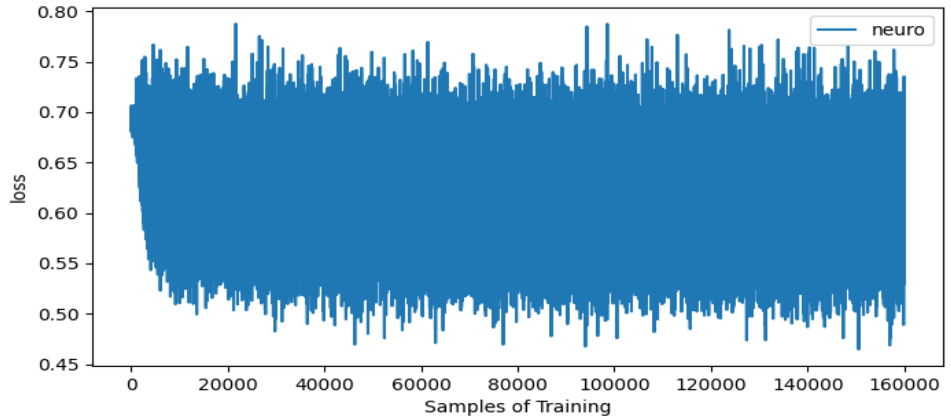
Входная m-грамма	Вероятность «0»	Вероятность «1»
«0000000000»	0.25	0.75
«0000000001»	0.75	0.25
…	…	…
«1111111111»	0.25	0.75

Таблица переходных вероятностей марковского источника

Марковский источник и FNN

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
FNN с 3мя скрытыми слоями [80,80,60]	Функция потерь - BCELoss	0.68575875	0.686425
	Функция потерь - KLDivLoss	0.6857275	0.68633
	Функция потерь - CrossEntropyLoss	0.68577625	0.686435
	Функция потерь - MSELoss	0.68574125	0.686405

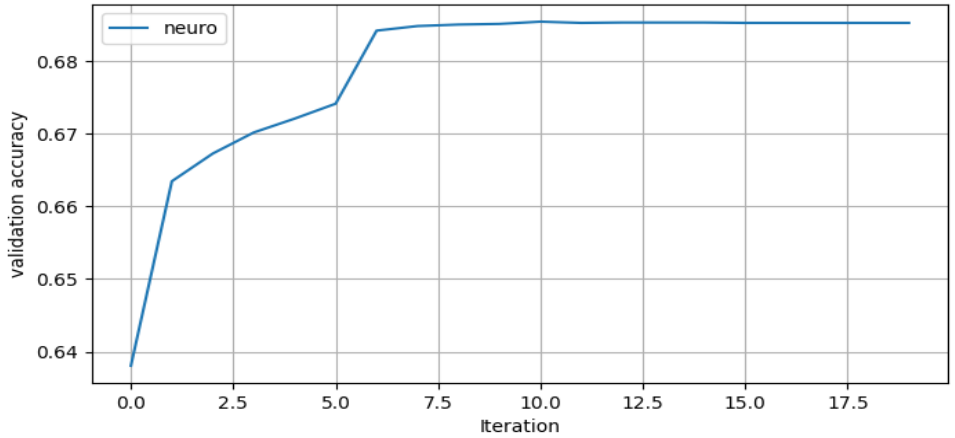


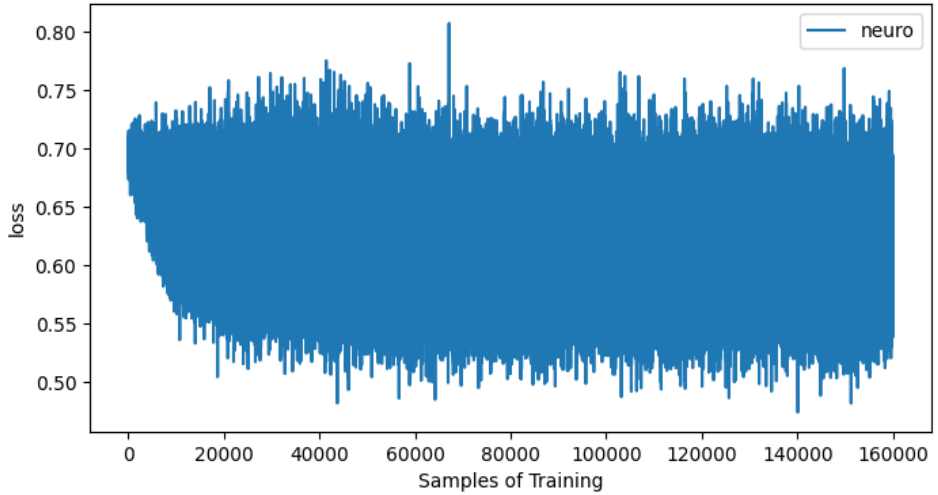


FNN с функцией CrossEntropyLoss

Марковский источник и ResNet

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
ResNet	5 скрытых слоев по 80 нейронов в каждом	0.68477	0.68529





Марковский источник и RNN

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
RNN (один слой, размерность скрытых признаков 40)	однонаправленная	0.68451125	0.684945
	двунаправленная	0.683895	0.68441

Марковский источник и CPN

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
CPN с метрикой cosine	размерность карты: 35x35	0.6568545	0.645715
CPN с метрикой euclidean	размерность карты: 35x35	0.658763	0.653325
	размерность карты: 50x50	0.6801	0.66928
CPN с метрикой manhattan	размерность карты: 35x35	0.6408	0.625425

Марковский источник глубины 10. Выборка без повторений и противоречий

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
FNN	• 3 скрытых слоям [80,80,60] • Функция потерь: CrossEntropyLoss	0.994174	0.68648
RNN	• один слой • размерность признаков 40 • однонаправленная	0.99996	0.68648
CPN	• размерность карты: 35x35 • метрика: euclidean	0.99973	0.61949

Периодический источник с периодом 20

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
FNN	Один слой из 100 нейронов	0.99997875	0.99998
RNN	• один слой • Размерность скрытых признаков 10	0.99997875	0.99998
CPN	• размерность карты: 10x10 • метрика евклидова	0.85	0.8499969
	• размерность карты: 20x20 • метрика евклидова	1.0	1.0

Линейный регистр сдвига с примитивным многочленом степени 40

Архитектура	Важные параметры	Точность на тренировочной выборке	Точность на тестовой выборке
FNN	• 3 скрытых слоям [80,80,60] • Функция потерь: CrossEntropyLoss	0.49103	0.50032
RNN	• один слой • размерность признаков 40 • однонаправленная	0.50194	0.506
CPN	• размерность карты: 35x35 • метрика: euclidean	0.49473	0.4949

Github

https://github.com/Respman/NIST_SP_800-90B_and_neural_nets

Заключение

- Рассмотренные архитектуры нейронных сетей будут работать как предикторы для случайных источников с конечной памятью и неприменимы для исследования алгоритмических ДСЧ
- Полученные результаты могут быть полезными при дальнейших исследованиях ФДСЧ

Вопросы

???

Контактная информация

- **Электронная почта:**
  Andrey.Sergeev@sfblaboratory.ru