

University of Warsaw  
Faculty of Mathematics, Informatics and Mechanics

**Marta Nowakowska**

Student no. 385914

# Time series prediction using self-attention models

Master's thesis  
in COMPUTER SCIENCE

Supervisor:

**dr Piotr Miłoś**

Institute of Mathematics, Polish Academy of Sciences

Co-supervisor:

**dr hab. Marek Cygan, prof. UW**

University of Warsaw

Warsaw, August 2022



## **Abstract**

TODO abstract

## **Keywords**

time series, machine learning, self-attention model

## **Thesis domain (Socrates-Erasmus subject area codes)**

11.4 Sztuczna inteligencja

## **Subject classification**

D. Software

D.127. Blabalgorithms

D.127.6. Numerical blabalysis

## **Tytuł pracy w języku polskim**

Prognozowanie szeregów czasowych przy użyciu modeli z uwagą



# Contents

<b>1. Introduction</b>	5
1.1. Time series and their importance	5
1.2. Problem statement	5
1.3. Contributions	6
<b>2. Podstawowe pojęcia</b>	7
2.1. Definicje	7
2.2. Blabalizator różnicowy	7
<b>3. Wcześniejsze implementacje blabalizatora różnicowego</b>	9
3.1. Podejście wprost	9
3.2. Metody wykorzystujące teorię Głombaskiego	9
3.3. Metody wykorzystujące własności fektorów $\sigma$	9
<b>4. Teoria fektorów <math>\sigma</math>-<math>\rho</math></b>	11
<b>5. Dokumentacja użytkowa i opis implementacji</b>	13
<b>6. Podsumowanie</b>	15
6.1. Perspektywy wykorzystania w przemyśle	15
<b>A. Główna pętla programu zapisana w języku TōFoo</b>	17
<b>B. Przykładowe dane wejściowe algorytmu</b>	19
<b>C. Przykładowe wyniki blabalizy (ze współczynnikami <math>\sigma</math>-<math>\rho</math>)</b>	21
<b>Bibliografia</b>	23



# Chapter 1

## Introduction

### 1.1. Time series and their importance

Time series is a sequence of data points measured across time with a fixed frequency. It is one of the most widespread type of data, making its forecasting incredibly useful for vast industries and research areas. Among others they are being used for stock market analysis, weather prediction and resource allocation planning.

It is not hard to imagine that this wide usage brought time series a lot of attention in the past. One of the most widely known method for their forecasting is ARIMA (autoregressive integrated moving average). Despite its popularity, it comes with some disadvantages, like the necessity to fit it separately for every similar time series and for the user to have some expertise in the parameter selection procedure.

This brings us to the next era of the time series forecasting: deep neural networks. The achievements in the natural language processing (NLP) field, including RNN and Transformer self-attention models have almost directly transferred to time series, due to a strikingly similar nature of both problems. However, RNNs and their successors (LSTM[2], GRU[3]) have difficulty with long sequences[4], causing them to struggle with time series that can contain long-term dependencies.

Transformer's[5] self-attention has revolutionized the world of long sequence processing, making it possible for the model to deal with larger-scale forecasting. However, there are several problems with integrating it directly for the time series forecasting usage: its local context-blindness[1] and large space-complexity growing quadratically with the sequence length. One solution was proposed in [1] by introducing convolutional self-attention and LogSparse Transformer.

### 1.2. Problem statement

Transformer proposed in [1] is centered around the idea of every point in a time series being drawn from some Gaussian distribution. Thus, the model's goal is to predict the parameters of these distributions, ensuring that each prediction is partially random and introducing Gaussian noise to it.

In reality, it's not always the case that every time series' noise comes from a Gaussian distribution. It would be more useful if we could model any continuous distribution instead. One way to approximate it is by using a discrete distribution with categories that are granular enough. Obviously, this comes with a disadvantage of being able to predict only a discretely limited value, which could cost some accuracy. However, the expectation is that this problem

can be solved easily by just using a sufficiently large number of categories.

The goal of our project was using the Transformer model to predict time series assuming that each point in the sequences belongs to a categorical distribution.

### **1.3. Contributions**

Work described in this thesis has been done via a collaboration with scientists from Institute of Mathematics of Polish Academy of Sciences.

My most important contributions are:

1. Gaussian distribution-based Transformer
2. Experiments with the Gaussian distribution-based Transformer
3. Base implementation of the dataset feeder
4. Time features as in [1] (on the data feeding side)



## Chapter 2

# Podstawowe pojęcia

Pojęciem pierwotnym blabalii fetorycznej jest *blaba*. Blabaliści nie podają jego definicji, mówiąc za Ciach-Pfe t-ām Kûn (fooistyczny mędrzec, XIX w. p.n.e.):

Blaba, który jest blaba, nie jest prawdziwym blaba.

tłum. z chińskiego Robert Blarbarucki

### 2.1. Definicje

Oto dwie definicje wprowadzające podstawowe pojęcia blabalii fetorycznej:

**Definicja 2.1.1** *Silny, zwarty i gotowy fetor bazowy nazwiemy skupieniem.*

**Definicja 2.1.2** *Fetorem nazwiemy skupienie blaba spełniające następujący aksjomat reperkusatywności:*

$$\forall \mathcal{X} \in Z(t) \exists \pi \subseteq \oint_{\Omega^2} \kappa \leftrightarrow 42$$

### 2.2. Blabalizator różnicowy

Teoretycy blabalii (zob. np. pracę [? ]) zadowalają się niekonstruktywnym opisem natury fetorów.

Podstawowym narzędziem blabalii empirycznej jest blabalizator różnicowy. Przyrząd ten pozwala w sposób przybliżony uzyskać współczynniki rozkładu Głombaskiego dla fetorów bazowych i harmoniczných. Praktyczne znaczenie tego procesu jest oczywiste: korzystając z reperkusatywności pozwala on przejść do przestrzeni  $\Lambda^\nabla$ , a tym samym znaleźć retroizotonalne współczynniki semi-quasi-celibatu dla klatek Rozkoszy (zob. [? ]).

Klasyczne algorytmy dla blabalizatora różnicowego wykorzystują:

1. dualizm falowo-korpuskularny, a w szczególności
  - (a) korpuskularną naturę fetorów,
  - (b) falową naturę blaba,
  - (c) falowo-korpuskularną naturę gryzmołów;
2. postępującą gryzmozolizację poszczególnych dziedzin nauki, w szczególności badań systemowych i rozcieńczonych;

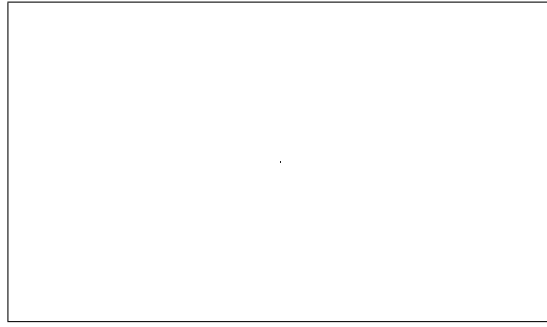


Figure 2.1: Artystyczna wizja blaba w obrazie węgierskiego artysty Josipa A. Rozkoszy pt. „Blaba”

3. dynamizm fazowy stetryczenia parajonizacyjnego;

4. wreszcie tradycyjne opozycje:

- duch — bakteria,
- mieć — chcieć,
- myśl — owsianka,
- parafina — durszlak<sup>1</sup>,
- logos — termos

z właściwym im przedziwym dynamizmem.

---

<sup>1</sup>Więcej o tym przypadku — patrz prace Gryzybór-Głombaskiego i innych teoretyków nurtu teoretyczno-praktycznego badań w Instytucie Podstawowych Problemów Blabalii w Fifie.

## Chapter 3

# Wcześniejsze implementacje blabalizatora różnicowego

### 3.1. Podejście wprost

Najprostszym sposobem wykonania blabalizy jest siłowe przeszukanie całej przestrzeni rozwiązań. Jednak, jak łatwo wyliczyć, rozmiar przestrzeni rozwiązań rośnie wykładniczo z liczbą fektorów bazowych. Tak więc przegląd wszystkich rozwiązań sprawdza się jedynie dla bardzo prostych przestrzeni lamblialnych. Oznacza to, że taka metoda ma niewielkie znaczenie praktyczne — w typowym przypadku z życia trzeba rozważać przestrzenie lamblialne wymiaru rzędu 1000.

W literaturze można znaleźć kilka prób opracowania heurystyk dla problemu blabalizy (por. [? ]). Korzystając z heurystyk daje się z pewnym trudem dokonać blabalizy w przestrzeni o np. 500 fektorach bazowych. Należy jednak pamiętać, że heurystyka nie daje gwarancji znalezienia najlepszego rozwiązania. Fifak w pracy [? ] podaje, jak dla dowolnie zadanej funkcji oceniającej skonstruować dane, dla których rozwiązanie wygenerowane przez algorytm heurystyczny jest dowolnie odległe od rzeczywistego.

### 3.2. Metody wykorzystujące teorię Głombaskiego

Teoria Głombaskiego (zob. [? ]) dostarcza eleganckiego narzędzia opisu przejścia do przestrzeni  $\Lambda^\nabla$ . Wystarczy mianowicie przedstawić fetyory bazowe wyjściowej przestrzeni lamblialnej w nieskończonej bazie tak zwanych wyższych aromatów. (Formalną definicję tego pojęcia przedstawię w rozdziale poświęconym teorii Fifaka). Podstawową cechą wyższych aromatów jest ulotność. To zaś oznacza, że odpowiednio dobierając współczynniki przejścia do przestrzeni wyższych aromatów można zagwarantować dowolną z góry zadaną dokładność przybliżonego rozwiązania problemu blabalizy.

Oczywiście ze względu na nieskończony wymiar przestrzeni wyższych aromatów koszt poszukiwania współczynników blabalizy jest liniowy ze względu na wymiar wyjściowej przestrzeni lamblialnej.

### 3.3. Metody wykorzystujące własności fektorów $\sigma$

Najchętniej wykorzystywaną przestrzenią wyższych aromatów jest przestrzeń fektorów  $\sigma$ . Fetyory  $\sigma$  dają szczególnie prostą bazę podprzestrzeni widłowej. Wiąże się to z faktem, że w tym przypadku fetyory harmoniczne wyższych rzędów są pomijalne (rzędu  $2^{-t^3}$ , gdzie  $t$  jest wymiarem wyjściowej przestrzeni lamblialnej).

Niestety z fektorami  $\sigma$  wiąże się też przykre ograniczenie: można wykazać (zob. [?, s. 374]), że dla dowolnie dobranej bazy w podprzestrzeni widłowej istnieje ograniczenie dolne w metryce sierpa na odległość rzutu dokładnego rozwiązania problemu blabalizy na podprzestrzeń widłową. Ponieważ rzut ten stanowi najlepsze przybliżone rozwiązanie, jakie można osiągnąć nie naruszając aksjomatu reperkusatywności, więc istnieje pewien nieprzekraczalny próg dokładności dla blabalizy wykonanej przez przejście do przestrzeni fektorów  $\sigma$ . Wartość retroinicjalną tego progu nazywa się *reziduum blabycznym*.

## Chapter 4

# Teoria fetorów $\sigma$ - $\rho$

Głównym odkryciem Fifaka jest, że fetor suprakowariantny może gryzmolizować dowolny ideał w podprzestrzeni widłowej przestrzeni lamblialnej funkcji Rozkoszy.

Udowodnienie tego faktu wymagało wykorzystania twierdzeń pochodzących z kilku niezależnych teorii matematycznych (zob. na przykład: [? ? ? ? ? ? ]). Jednym z filarów dowodu jest teoria odwzorowań owalnych Leukocyta (zob. [? ]).

Znaczenie twierdzenia Fifaka dla problemu blabalizy polega na tym, że znając retroizotonalne współczynniki dla klatek Rozkoszy można przeprowadzić fetory bazowe na dwie nieskończone bazy fetorów  $\sigma$  w przestrzeni  $K_7$  i fetorów  $\rho$  w odpowiedniej quasi-quasi-przestrzeni równoległej (zob. [? ]). Zasadnicza różnica w stosunku do innych metod blabalizy polega na tym, że przedstawienie to jest dokładne.



## Chapter 5

# Dokumentacja użytkowa i opis implementacji

Program przygotowany dla systemu operacyjnego M\$ Windows uruchamia się energicznym dwumlaskiem na jego ikonce w folderze `\\FIDO\F00\BLABA`. Następnie kolistym ruchem ręki należy naprowadzić kursor na menu **Blabaliza** i uaktywnić pozycję **Otwórz plik**. Po wybraniu pliku i zatwierdzeniu wyboru przyciskiem **OK** rozpocznie się proces blabalizy. Wyniki zostaną zapisane w pliku o nazwie `99-1a.tx.43` w bieżącym folderze.





## Chapter 6

# Podsumowanie

W pracy przedstawiono pierwszą efektywną implementację blabalizatora różnicowego. Umiejętność wykonania blabalizy numerycznej dla danych „z życia” stanowi dla blabalii fetorycznej podobny przełom, jak dla innych dziedzin wiedzy stanowiło ogłoszenie teorii Mikołaja Kopernika i Gryzybór Głombaskiego. Z pewnością w rozpoczynającym się XXI wieku będziemy obserwować rozkwit blabalii fetorycznej.

Trudno przewidzieć wszystkie nowe możliwości, ale te co bardziej oczywiste można wskazać już teraz. Są to:

- degryzmolizacja wieńców telecentrycznych,
- realizacja zimnej reakcji lambliarnej,
- loty celulityczne,
- dokładne obliczenie wieku Wszechświata.

### 6.1. Perspektywy wykorzystania w przemyśle

Ze względu na znaczenie strategiczne wyników pracy ten punkt uległ utajnieniu.



## Appendix A

# Główna pętla programu zapisana w języku TōFoo

```
[[foo]{,}[a3,((([,){[[]]})),
  [1; [{,13},[[[11],11],231]]].
  [13;[!xz]].
  [42;[{,x},[[2],{'a'},14]]].
  [br;[XQ*10]].
), 2q, for, [1,]2, [..].[7]{x}],[(,([1{{123},},,;.112]],
  else 42;
  . 'b'.. '9', [[13141],{13414}], 11),
[1; [[134,sigma],22]].
[2; [[rho,-],11]].
)[14].
), {1234}],]. [map [cc], 1, 22]. [rho x 1]. {22; [22]},
  dd.
[11; sigma].
  ss.4.c.q.42.b.ll.ls.chmod.aux.rm.foo;
[112.34; rho];
  001110101010101010101010101010101011111010010
[22%f4].
cq. rep. else 7;
]. hlt
```



## Appendix B

### Przykładowe dane wejściowe algorytmu

$\alpha$	$\beta$	$\gamma$
901384	13784	1341
68746546	13498	09165
918324719	1789	1310
9089	91032874	1873
1	9187	19032874193
90143	01938	0193284
309132	-1349	-149089088
0202122	1234132	918324098
11234	-109234	1934



## Appendix C

### Przykładowe wyniki blabalizy (ze współczynnikami $\sigma$ - $\rho$ )

	Współczynniki Głombaskiego	$\rho$	$\sigma$	$\sigma$ - $\rho$
$\gamma_0$	1,331	2,01	13,42	0,01
$\gamma_1$	1,331	113,01	13,42	0,01
$\gamma_2$	1,332	0,01	13,42	0,01
$\gamma_3$	1,331	51,01	13,42	0,01
$\gamma_4$	1,332	3165,01	13,42	0,01
$\gamma_5$	1,331	1,01	13,42	0,01
$\gamma_6$	1,330	0,01	13,42	0,01
$\gamma_7$	1,331	16435,01	13,42	0,01
$\gamma_8$	1,332	865336,01	13,42	0,01
$\gamma_9$	1,331	34,01	13,42	0,01
$\gamma_{11}$	1,331	8913,01	13,42	0,01
$\gamma_{12}$	1,331	13,01	13,42	0,01
$\gamma_{13}$	1,334	789,01	13,42	0,01
$\gamma_{14}$	1,331	4897453,01	13,42	0,01
$\gamma_{15}$	1,329	783591,01	13,42	0,01





# Bibliography

- [1] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyu Zhou, Wenhui Chen, Yu-Xiang Wang and Xifeng Yan. *Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting*. arXiv:1907.00235, 2019.
- [2] Sepp Hochreiter and Jürgen Schmidhuber. *Long short-term memory*. Neural computation, 9(8):1735–1780, 1997.
- [3] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. *On the properties of neural machine translation: Encoder-decoder approaches*. arXiv preprint arXiv:1409.1259, 2014.
- [4] Urvashi Khandelwal, He He, Peng Qi, and Dan Jurafsky. *Sharp nearby, fuzzy far away: How neural language models use context*. arXiv preprint arXiv:1805.04623, 2018.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. *Attention is all you need*. NIPS 2017.