ISSN 2579-2784 (Print)
ISSN 2538-2788 (Online)

MATHEMATICAL PROBLEMS OF COMPUTER SCIENCE

LX

Yerevan
2023

Հայաստանի Հանրապետության Գիտությունների ազգային ակադեմիայի
Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ

Институт проблем информатики и автоматизации Национальной академии наук Республики Армения

Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ

Математические проблемы компьютерных наук

Mathematical Problems of Computer Science

LX

ՀՐԱՏԱՐԱԿՎԱԾ Է ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏԻ ԿՈՂՄԻՑ
ОПУБЛИКОВАНО ИНСТИТУТОМ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА
PUBLISHED BY THE INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF NAS RA

ԵՐԵՎԱՆ 2023 YEREVAN

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ, LX

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ պարբերականը հրատարակվում է տարեկան երկու անգամ ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի (ԻԱՊԻ) կողմից: Այն ընդգրկում է տեսական և կիրառական մաթեմատիկայի, ինֆորմատիկայի և հաշվողական տեխնիկայի ժամանակակից ուղղությունները:

Այն ընդգրկված է Բարձրագույն որակավորման հանձնաժողովի ընդունելի ամսագրերի ցանկում:

Տպագրվում է Խմբագրական խորհրդի 2023թ. նոյեմբերի 23-ի N 23-11/1 նիստի որոշման հիման վրա

ISSN 2579-2784 (Print)
ISSN 2738-2788 (Online)

Журнал **Математические проблемы компьютерных наук** издается два раза в год Институтом проблем информатики и автоматизации НАН РА. Он охватывает современные направления теоретической и прикладной математики, информатики и вычислительной техники.

Он включен в список допустимых журналов Высшей квалификационной комиссии.

Печатается на основании решения N 23-11/1 заседания
Редакционного совета от 23 ноября 2023г.

ISSN 2579-2784 (Print)
ISSN 2738-2788 (Online)

Mathematical Problems of Computer Science, LX

The periodical **Mathematical Problems of Computer Science** is published twice per year by the Institute for Informatics and Automation Problems of NAS RA. It covers modern directions of theoretical and applied mathematics, informatics and computer science.

It is included in the list of acceptable journals of the Higher Qualification Committee.

Printed on the basis of decision N 23-11/1 of the session of the Editorial Council dated November 23, 2023.

ISSN 2579-2784 (Print)
ISSN 2738-2788 (Online)

Պարբերականի սույն պրակը նվիրվում է տեխնիկական գիտությունների թեկնածու Սուրեն Բախշիի Ալավերդյանի հիշատակին:

Этот выпуск журнала посвящен памяти кандидата технических наук Сурена Бахшиевича Алавердяна.

The current volume is dedicated to the memory of Suren Bakhshi Alaverdyan, a Candidate of Technical Sciences.

Suren Alaverdyan (1952-2023)

S. Alaverdyan graduated from the Faculty of Applied Mathematics of Yerevan State University in 1973. He started working at the IIAP of NAS RA that same year. In 1988, he completed his candidate's thesis.

He worked on digital image processing, designing and implementing application programs. Under his leadership, methods were developed for the recognition of printed Armenian texts, within the framework of which a system for recognizing printed texts was created. Another software package created under his leadership provides an opportunity to restore the overall picture from cartographic fragments. His long-term scientific activity is also presented in many scientific articles. Another software package was developed under his leadership, which provides an opportunity to restore the overall picture from cartographic fragments. Additionally, he has contributed significantly to the scientific community through numerous published scientific articles that showcase his long-term scientific activity.

S. Alaverdyan was highly skilled in the field of digital image processing and had a vast knowledge of various methods used in the field. He was a dedicated researcher and programmer who contributed significantly to the development of the institute's young employees by imparting his knowledge and skills to them.

He always earned the love and respect of the collective, and they will hold him in their thoughts and memories for a long time.

CONTENTS

Mathematical Problems of Computer Science 60, 9–16, 2023.
doi: 10.51408/1963-0104

UDC 519.6

Image Reconstruction Using the *sinc* Kernel Function

Suren B. Alaverdyan

Institute for Informatics and Automation Problems of NAS RA,Yerevan, Armenia
e-mail: souren@iiap.sci.am

Abstract

This study is devoted to address the challenge of solving ill-posed integral equations for image restoration. The integral equation is widely recognized as an ill-posed problem [1]. Our study demonstrates that utilizing a two- dimensional function as a kernel function in the integral equation leads to stable solutions, by establishing a consistent dependence between the solutions and the input data (images).

We were able to obtain solutions for the integral equation without employing a regularization process, which significantly reduces the duration of the calculation process.
Keywords: Integral equation, Correct, Kernel function.
Article info: Received 16 March 2023; sent for review 17 April 2023; accepted 18 May 2023.

1. Introduction

There are many publications on the subject of image restoration, and ongoing research indicates its continued importance in the field.

Numerous methods have been developed for image recovery, including some well-known examples such as Wiener's method of filtration [2], Tikhonov's regularization method for solving numerical solutions to ill-posed double Fredholm integral equations of the first kind [1, 3], and a family of methods that utilize the bundle theorem's result. They are known as blind deconvolution. Additionally, some methods rely on spectral analysisto restore the image by altering the spectrum values corresponding to low and high frequencies.

It is worth noting that in solving the image restoration problem, additional challenges may arise, such as estimating the degree of image distortion, determining the radius of the scattering function, and finding the image sharpness coefficient or estimation.

The precision coefficient can be used as a characteristic or be one of the characteristics of the completion of the iteration process.

Since images can have both low-frequency and high-frequency noises simultaneously, the filtering problem becomes quite delicate. Implementing one filtering method can adversely affect another one, and vice versa. To address this issue, spectral analysis can be used to simultaneously alternate the spectrum values in the low-frequency and high-frequency ranges.

Considering the importance of V. Kotelnikov's signal recovery theorem, which uses orthogonal basis functions $\eta\ y_n = sinc(x-n),\ n \in Z$, however, when using the two- dimensional $sinc(x-n, y-\eta) = sinc(x-n)sinc(y-\eta)$ function as a kernel function in the integral equation for image recovery, a fundamental question arises whether this approach can make it a well-posed problem or not. Implementation of the algorithm in the time domain has yielded positive results making this work worthwhile.

2. Integral Equation for Image Recovery

The integral equation for image reconstruction is a two-dimensional Fredholm integral equation of the first type (the unknown function is contained within the subintegral expression) and is expressed as follows:

$$\int_a^b \int_c^d k(x-\xi, y-\eta) f(\xi, \eta) d\xi d\eta = g(x, y), \tag{1}$$

where $a \leq x, \xi \leq b$, $c \leq y, \eta \leq d$.

The function $k(x-\xi, y-\eta)$ is called a kernel of the integral equation and is also known as a point dispersion (or spread) function, and the integral equation has a free term: $g(x, y)$, which is a given function (representing approximate data, images, etc.), $f(\xi, \eta)$ is an unknown function. This equation characterizes a variety of other physical processes, such as tomography and chemical-mechanical smoothing [4].

It is assumed that the function $k(x, y)$ is a quadratic integrable function:

$$\int_a^b \int_c^d |k(x, y)|^2 dxdy < \infty.$$

Here are some examples of kernel functions:

- Gaussian kernel: $k(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$;
- Distance-based kernel: $k(x, y) = \frac{1}{\sqrt{x^2+y^2}}$;
- sinc kernel: $k(x, y) = sinc(x, y) = sinc(x)sind(y)$.

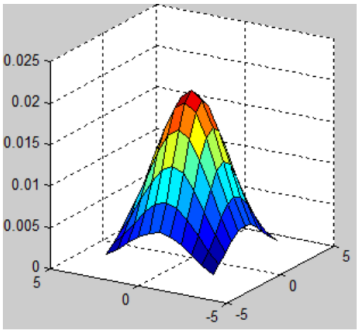

Fig. 1. Gaussian kernel.

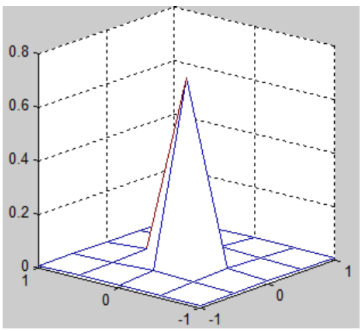

Fig. 2. Distance-based kernel.

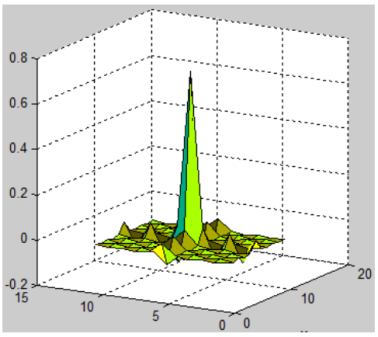

Fig. 3. sinc kernel.

3. $sinc(x)$ Function Properties

Let's discuss some properties that are significant in various fields of data processing. Most of these properties are used in the implementation of the algorithm.

$$\int_{-\infty}^{+\infty} sinc(x)dx = \pi.$$

The formula shows that the normalized $sinc(x)$ function takes the following form:

$$sinc(x) = \frac{sinc(\pi x)}{\pi x}.$$

The following relation is valid for integers:

$$sinc(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 : \end{cases} \tag{2}$$

The Fourier transform of $sinc(x)$ is called an orthogonal function, which has the following form:

$$\int_{-\infty}^{+\infty} sinc(x)e^{-2\pi i\omega x}dx = rect(\omega), \tag{3}$$

where the bounded rectangular function is expressed as follows:

$$rect(\omega) = \begin{cases} 1, & if \quad |\omega| \leq a, \\ 0, & if \quad |\omega| > a, \end{cases}$$

where $a = const$.

For practical applications, it is more convenient to represent the function $rect(\omega)$ in the following way:

$$rect(\omega) = \begin{cases} 1, & if \quad 0 \leq \omega \leq 1, \\ 0, & out \ \ of \ \ range. \end{cases}$$

The frequency value of $\omega_0 = 0.5$ is commonly referred to as a center frequency. The frequency range of $\omega_0 < 0.5$ is considered as a low-pass-frequency range while $\omega_0 > 0.5$ is referred to as the high-frequency range. It is worth noting that at the value of $\omega_0 = 0.5$, the filtered function retains its original value.

We can show that the bounded delta function with the integral representation is reduced to the $sinc(x)$ function:

$$\delta(\omega) = \int_{-a}^{a} e^{-2i\omega t}dt = 2\int_{0}^{a} cos(\omega a)dt = \frac{2asin(\omega a)}{\omega a} = 2asinc(\omega a). \tag{4}$$

The delta function has a filtering property. Let us now examine the graphs of the $sinc(x)$ function and its spectrum (see Fig. 4).

Signals with constant spectral density are called white noise, which contains the entire range of frequencies: $(0, \infty)$. By definition, the function $rect(\omega)$ is the spectral density of low- frequency $\omega < 5$ limited white noise. From (3) and (4) it follows that the $sinc$ function is the covariance function of low-frequency smooth noise.

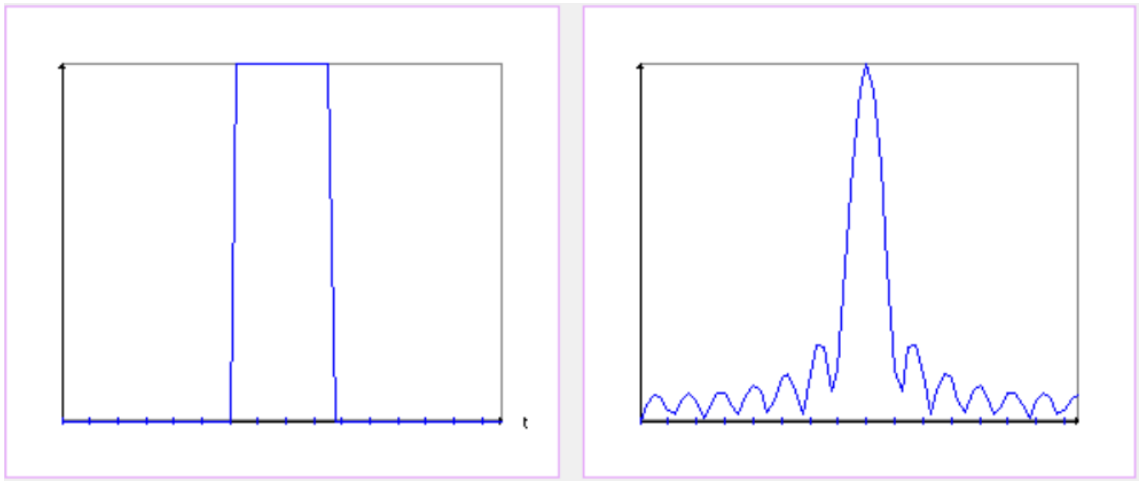

Fig. 4. Graphs of the $sinc(x)$ function and its amplitude spectrum.

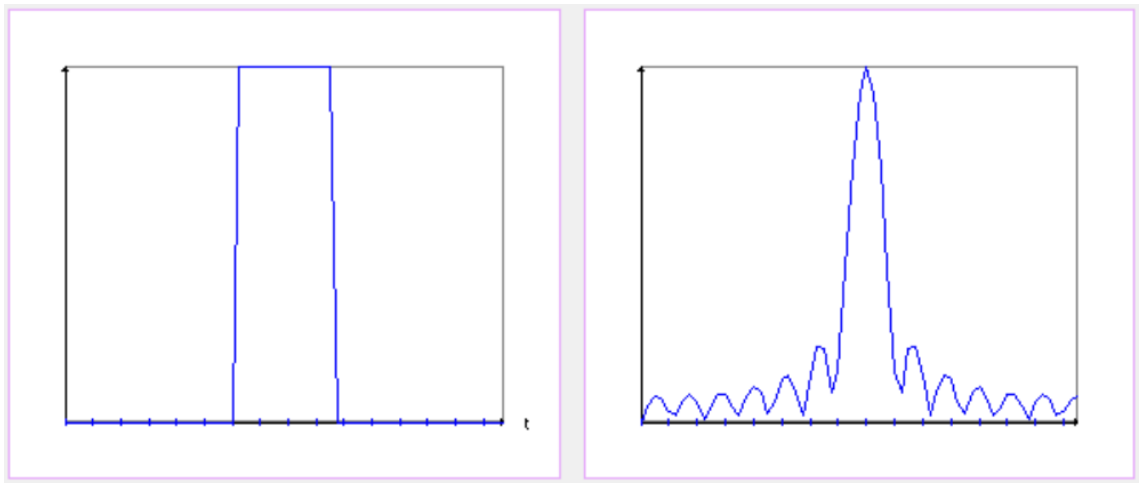

Fig. 5. Graphs of the $rect(x)$ function and its amplitude spectrum.

Formula (2) shows that for all values of n, the vectors $y_n = sinc(x - n),\ n \in Z$ form an orthonormal basis, which is used in signal recovery (V. Kotelnikov's theorem).

According to the theorem (V. Kotelnikov), any function $f(t)$, consisting of frequencies from 0 tof_c, can be transmitted continuously with any precision, using uniformly spaced samples taken at intervals of $1/(2f_c)$ seconds.

Any function $f(t)$ containing frequencies between 0 and f_c can be transmitted continuously with any level of precision using samples taken at intervals of $1/(2f_c)$ seconds.

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta) sinc\left[\frac{\pi}{\Delta}(t - k\Delta)\right], \quad 0 < \Delta \leq \frac{1}{2f_c}.$$

For a signal to be accurately restored, it needs to be broadcast at a frequency of more than twice its maximum frequency. For instance, audio signals are commonly broadcast at a frequency of 44,000 hertz, given that the highest frequency audible to humans is 20,000 hertz. Additionally, Formula (4) shows the equivalence of the **delta** function of the $sinc(x)$ function. In physics, problems involving the **delta** function are typically handled by using the $sinc(x)$ function within a small range during calculations.

4. The *sinc* Function and Image Reconstruction in the Time Domain

An image is a projection of reflected electromagnetic waves onto a receiver. The main characteristics of an image, such as its resolution (the number of points per unit area), illumination, color, and contrast. These characteristics are significantly different from those characteristics of the signal that created it, such as its frequency, amplitude, and phase. However, there is a commonality between them: both are the result of wave processes.

In images, frequency and phase have a hidden nature. It becomes an object of study after the determination of its spectrum. The mentioned generality allows us to assume that the image is also a signal and represent it as a linear combination of the two-dimensional function $sinc(x-\xi)sinc(y-\eta)$ and an unknown function.

In this case, the integral equation (1) will take the following form:

$$\int_a^b \int_c^d sinc(x-\xi, y-\eta) f(\xi,\eta) d\xi d\eta = g(x,y). \tag{5}$$

To implement *sinc* filtering in the time domain, we can present (5) in a discrete form, where the $sinc(x)$ function will appear in a normalized form:

$$sinc(x-\xi, y-\eta) f(\xi,\eta) = \frac{sin(2\pi\omega(x-\xi))}{2\pi\omega(x-\xi)} \times \frac{sin(2\pi\omega(y-\eta))}{2\pi\omega(y-\eta)}. \tag{6}$$

Let m, n be the number of splits in $[a,b]$ and $[c,d]$ intervals:

$$dy = \frac{b-a}{m},\ dx = \frac{d-c}{n}.$$

We can represent the functions $f(\xi,\eta)$, $g(x,y)$, $sinc(x)sinc(y)$ as matrices F, G and S, respectively.

The G matrix has dimensions $m \times n$, while the matrix corresponding to the $sinc(x)sinc(y)$ kernel function is a square matrix, the size of which is optional: $k = 2p+1$. This number is chosen as an odd value to ensure the symmetry of transformations and prevent data skewing. Typically, algorithms are implemented for values of k=3.5, as larger values dramatically increase computation time.

To represent the unknown matrix F from (5) in the time domain, we use the known matrix G and a discrete package (such as convolution) of the kernel function (6):

$$F = G * S,$$

where $*$ is the convolution operation symbol.

The dimensions of the F matrix are $(m+2p) \times (n+2p)$. The formula for the discrete package looks as follows:

$$f_{i,j} = \sum_{u=i-p}^{i+p} \sum_{v=j-p}^{j+p} g_{i,j} s_{u-i+p, v-j+p}, \quad i = p, p+1, ..., m-p; \ \ j = p+1, ..., n-p. \tag{7}$$

It is important to note that while implementing the algorithm, the dimensions of the F matrix may not be changed, but the indices u and v should be controlled to ensure that they remain within the ranges $[0,m)$ and $[0,n)$.

Here is an example of the result of filtering an image containing Gaussian noise using (7) (see Fig. 6).

a b c

Fig. 6. a- original image, b- containing noise, $\sigma = 2$, $\rho = 9$, c- recovered image.

5. *sinc* Function and Image Restoration in the Spectral Domain

Image restoration in the spectral domain is performed by finding a solution to the linear integral Equation (5). To find the unknown function, the two-dimensional Fourier transform and the bundle theorem (convolution theorem) are used.

The set of functions $f(x)$ and $g(x)$ of one variable is the following integral:

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(t)g(x-t)dt. \tag{8}$$

To understand the meaning of the package formula, let's consider a simple example that deals with the product of two polynomials. Consider two polynomials $P_1(x) = 2x - 1$; $P_2(x) = x^2 - 3x + 1$. We are required to find the product polynomial of these polynomials.

Let's write the coefficients of these polynomials in the following form, starting with the coefficients of the polynomial $P_1(x)$ in reverse order: $(-1, 2)$.

1)	$P_2(x)$:	1,	−3,	1	2)	1,	−3,	1	3)	1,	−3,	1	4)	1,	−3,	1
	$P_1(x)$:		−1,	2			− 1,	2			− 1,	2			− 1,	2

It can be observed that at each step, the coefficients of the polynomial $P_1(x)$ are shifted one step to the left. The coefficients of the first polynomial are multiplied by the coefficients in front of them and added up. As a result we get the following:

$$1)\ -1 \cdot 1 = -1;\quad 2)\ -1 \cdot (-3) + 1 * 2 = 5;\quad 3)\ -1 \cdot 1 + 2 * (-3) = -7;\quad 4)\ 2 \cdot 1 = 2.$$

We obtain the coefficients of the product polynomial in reverse order. Correcting them we will get the following numbers: 2, -7, 5, -1.

$$P_1(x)P_2(x) = 2x^3 - 7x^2 + 5x - 1.$$

This calculation helps to characterize the significance of the bounded integral in Formula (8). Using the Fourier transform, the product of two functions can be computed efficiently. If $\hat{F}$ and $\hat{G}$ are transforms of the of functions f and g, respectively, then the product is calculated by the following formula:

$$(f * g)(x) = \Phi^{-1}(\hat{F} \cdot \hat{G}),$$

where $(\cdot)\hat{F} \cdot \hat{G}$ is the product of the corresponding elements of the vectors, and Φ^{-1} is the inverse Fourier transform.

According to the bundle theorem, the right-hand side of Equation (5) is the two-dimensional package of functions $sinc(x-\xi, y-\eta)$ and $f(\xi,\eta)$, so it can be represented in the following form:

$$\Phi_2(f(\xi,\eta) \cdot sinc(x,y)) = \Phi_2(g(x,y)). \tag{9}$$

Φ_2 and Φ_2^{-1} are direct and inverse two-dimensional Fourier transforms.

Let us denote the Fourier transforms of the functions $f(\xi,\eta)$, $sinc(x,y)$, $g(x,y)$ as $\hat{F}$, $\hat{S}$ and $\hat{G}$, respectively. Since we are dealing with a linear equation and linear transformations, we can obtain the following expression from Formula (9):

$$\hat{F} \cdot \hat{S} = \hat{G}, \quad or \quad \hat{F} = \frac{\hat{G}}{\hat{S}},$$

where we find the unknown function:

$$f(x,y) = \Phi_2^{-1}\left(\frac{\hat{G}}{\hat{S}}\right).$$

Since image restoration is an iterative process, an estimate of the image quality is needed to determine when the process should stop. This estimate is typically based on the normalized image gradient and Laplacian, which characterize the contours of objects in the image.

Given an image A having $m \times n$ patches. The coefficient of sharpness is determined by the following formula:

$$s = \frac{1}{2mna_{max}}\left(\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial^2 a_{ij}}{\partial x^2} + \frac{\partial^2 a_{ij}}{\partial y^2}\right)^2} + \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial a_{ij}}{\partial x} + \frac{\partial a_{ij}}{\partial y}\right)^2}\right).$$

In the given example shown in Fig. 6, for the input image 6-a, which doesn't contain any noise, the image sharpness coefficient value is assumed to be $s = 0.656817$. Then, low-frequency Gaussian noise with *radiuse* $\sigma = 2$, $r = 9$ is then added to the image 6-b, and the sharpness coefficient is calculated again, resulting in $s = 0.254214$. After filtering in the time domain, the sharpness coefficient is improved to $s = 0.583647$, and after filtering in the spectral domain, the sharpness coefficient is further improved to $s = 0.570277$.

6. Conclusion

The paper presents a method for finding solutions to the integral equation for image restoration, which bypasses the regularization process of an incorrectly set integral equation. This is achieved by selecting an appropriate dispersion function (kernel) of the integral equation of image restoration. The bounded two-dimensional $sinc(x,y)$ function was chosen as the kernel.

References

[1] Ж. Адамар, *Задача Коши для линейных уравнений с частными производными гиперболического типа*, М. Наука, 1978.

[2] Н. Ахмед и К. Рао, *Ортогональные преобразования при обработке цифровых сигналов*, Москва, Связь, 1980.

[3] А. Н. Тихонов, «О регуляризации некорректно поставленных задач», *ДАН СССР*, т. 153, н. 1, сс. 49 - 52, 1963.

[4] R. Ghulghazaryan, S. Alaverdyan and D. Piliposyan, "Accuarate pressure calculation method for CMP modeling using Fourier analysis", *Reviesed Selected papers, Computer Science and Information Technologies (CSIT)*, doi: 10.1109/CSITechnol.2019.8895113, pp. 43-46, 2019.

Պատկերների վերականգնումը *sinc* միջուկի ֆունկցիայի միջոցով

Սուրեն Բ. Ալավերդյան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
e-mail: souren@iiap.sci.am

Ամփոփում

Աշխատանքում ներկայացվել է պատկերների վերականգնման ինտեգրալ հավասարման լուծումներ գտնելու մեթոդ, որը շրջանցում է ոչ կոռեկտ դրված ինտեգրալ հավասարման ռեգուլյարիզացիայի պրոցեսը: Այս հանգամանքը պայմանավորված է պատկերների վերականգնման ինտեգրալ հավասարման ցրման ֆունկցիայի (կորիզ, kernel) ընտրությամբ: Որպես կորիզ ընտրվել է սահմանափակ երկչափ $sinc(x, y)$ ֆունկցիան:

Բանալի բառեր՝ Ինտեգրալ հավասարում, կոռեկտ, միջուկ, ֆունկցիա:

Реконструкция изображения с использованием функции ядра *sinc*

Сурен Б. Алавердян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
e-mail: souren@iiap.sci.am

Аннотация

В работе представлен метод нахождения решения интегрального уравнения восстановления изображения, который позволяет обойти процесс регуляризации некорректно заданного интегрального уравнения. Это обстоятельство обусловлено выбором функции ядра интегрального уравнения восстановления изображения. В качестве ядра была выбрана ограниченная двумерная функция $sinc(x, y)$.

Ключевые слова: Интегральное уравнение, корректный, ядро, функция.

Mathematical Problems of Computer Science 60, 17–26, 2023.
doi: 10.51408/1963-0105

UDC 004.8

Promoting Origination of Dynamicity of Non-Cellular Cognizers

Edward M. Pogossian

Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
e-mail: epogossi@aua.am

Abstract

Dynamic realities exist at the most basic level of elementary particles, which, according to quantum field theory emerge as excitations of fundamental quantum fields. At the same time, the nuclei of cognizers – doers, and their modes -1/2place classifiers and energizers, are also types of dynamic realities. Trying to trace the origin of the dynamicity of doers to the dynamics of particles and fields would help enlighten the origination of classifiers in nature.

As a footstep to a positive answer to this question, we provide cases of such interpretation of dynamicity concluding by some hints to generalize the cases.

Keywords: Modeling, Doers, Energizers, Classifiers, Cognizing, Fundamentals, Dynamicity.

Article info: Received 25 September 2023; sent for review 27 September 2023; received in revised form 6 November 2023; accepted 13 November 2023.

Acknowledgment: The author expresses his gratitude to Levon Pogosian for productive consultations and corrections.

1. Introduction

1.1. Questioning the origination of non-cellular cognizers, i.e., their nuclei, *classifiers*, and then *energizers* [1,2], inevitably refers to the nature and origination of the most reliable, fundamental categories of the universe U* such as fields, energy, particles, atoms and their compounds.
We can reasonably expect that the explanation of non-cellular classifiers and energizers by reliable fundamentals and regularities of their alliancing could be supportive in revealing their origination, especially, when their appearance can be regularized.
1.1.1. These expectations rely, at least, on the following premises.

At first, in nature, durable classifiers can be identified as absorbers (it seems even recurrent), along with widespread single classifiers as atoms that identify other atoms forming certain outputs, molecules, but unfortunately, performing this only lonely, solitarily, i.e., not recurrently.

1.1.2. Recall also that 1/2-place classifiers and energizers are one of the modes of a type of dynamic realities - *doers*, defined as realities having input-output parts and for realities at the input parts elaborating certain output ones or remain passive [1,2]**.**

1.1.3. Consequently, it is reasonable to assume that the existence or origination of recurrent 1/2-place classifiers and, generally doers, are not excluded in nature.

Then, recalling that dynamic realities are one of the fundamentals of quantum field theory (QFT), it is also reasonable to assume that enriching the dynamicity of doers with those in QFT could be supportive in interpreting classifiers by fundamentals of QFT and might enlighten the origination of classifiers and energizers in nature.

1.2. Origination of recurrent classifiers, directly or not, is questioned, particularly, in [3,4,5,6], as well as in [7].

1.2.1. All molecules in hypercycles are linked so that each of them catalyzes the creation of its successor, with the last molecule catalyzing the first one. In such a manner, the cycle reinforces itself. Furthermore, it is assumed that each molecule is additionally a subject for self-replication, hypercycles could originate naturally, and the incorporation of new molecules can extend them [3,4].

So, an exciting challenge is to link the recurrence of hypercycles with recurrent classifiers of cognizers.

1.2.2. Communication, we assume, necessarily requires regular identification of the IDs of correspondents, and thus, their possession of recurrent classifiers of IDs.

1.2.2.1. Indeed, communication between corespondents r and r' assumes they are nominated.

Then, as it is refined in [1], nominated corespondents inevitable own recurrent classifiers of ID's. Namely, communication of r,r' presumes they' contain classifiers that identify IDs of r,r', followed by their processing.

1.2.2.2. Consequently, in such interpretation communicating living molecules, argued in [5], unavoidably have to control and process certain recurrent classifiers.

Thus, statement in [5] (and, generally, in ontogenesis by Pierce [6]) that "…language is a general principle of Nature" and the statement that "recurrent classifiers are primordial" can be interpreted as equal.

1.3. In what follows we question, whether doers can be interpreted as models of dynamicity by QFT and as a positive footstep provide a variety of cases of modeling dynamicity by doers.

In consequent chapters we detail the above intentions as follows.

At first, we position ourselves in the approvals of the hypothesis, followed by takeaways from originations by QFT focused on acknowledged originations and formations there.

Then, asking whether doers can be interpreted as models of dynamicity by QFT, we provide a variety of cases of such interpretation and summarize them.

2. Focusing Approval of Hypothesis

2.1. Questioning origination of non-cellular cognizers refers, first of all, to the approval of origination of the *nuclei of the roots of cognizers* (nrcogs) – a variety of *1/2place classifiers* either symbolic or not, then, *energizers*.**,** based on the category of energy and its processing by constituents of classification and accumulation of energy, as well as overall compartmenting and reproduction.

Correspondingly, the answer to the questioning is reduced to a reliable explanation of constituents of nrcogs based on a contemporary theory of physicists on energy and fundamentals, i.e., on QFT.

Thus, the answer should provide reliable cause-effect chains between classifiers of fundamentals of QFT and those of nrcogs.
These chains inevitably should refer to the reliability of classifiers of fundamentals such as energy, fields, particles, and atoms, as well as to the reliability of relationships chaining these classifiers with the classifiers of nrcogs.
Thus, to advance in a reliable explanation of the origination of nrcogs, we need to address the reliability scales of these classifiers with respect to the utilities they identify.
2.2. For this aim, recall that the classifiers of community members are mainly inherited or acquired from genomes and cultures of communities while they could be revealed and contributed to the cultures in a lifetime.
Classifiers allow community members to identify favoring or damaging realities, i.e., to identify utilities of realities, to do, to act with them properly.
2.2.1. Recall also that the reliability of classifiers Cl we interpret as a measure of difference between the utilities promised by Cl for identified by Cl samples and ones actually provided to the members by these samples [2].
2.2.2. Our basic, primary do-classifiers dCl identify realities with certain procedures straightly, while complex, system-classifiers sCl identify them as a result of processing the constituent of sCl that altogether comprise the meanings of sCl and are associated with certain communicates [1].
Communicates (cms) can be IDs of meanings, compositions of IDs of do-classifiers of meanings, as well as compositions of samples of input domains (indoms) of classifiers.
2.2.3. Communicating members of communities reciprocally explain and understand the meanings of their cms to coordinate their efforts in attaining or preserving common utilities.
Communications are effective because the meanings of cms of the members are nearly equal. This equality is caused either by the fact that meanings of the members were directly acquired from the same communities or were revealed inductively by approximately equal means for all members, caused, first of all, by the commonality of their genomes.
2.3. Let us underline also that equal understanding of some cms *c* in communities C means only that members of C have approximately equal meanings *mc* on *c*, while the degrees of cohesion of utilities of realities identified by *mc* classifiers, i.e., identification of realities with prescribed to *mc* classifiers utilities, or quality of *mc* classifiers, can vary greatly.
2.4. Depending on the amount of experience, not always suitable goals and some other reasons argued in [2], classifiers, in general, can be far from perfectness, i.e., not sufficiently reliable in identifying the utilities associated with them.
Indeed, all classifiers are formed with a variety of types and amounts of experience, while massively experienced classifiers are stated as *postulates*.
Postulates combined with classifiers inferred from postulates by some rules/relationships (say by *cause-effect* one or its *modus pones* abstraction in logics) and possibly together with some not inferred assertions, hypotheses, comprise *theories,* where inference rules themselves are also based on experience also approved massively.
2.4.1. Resuming, it can be assumed that the more dense and reliable the cause-effect chains linking classifiers of target nrcogs to those with already acknowledged degrees of reliability in theories, the closer the reliability of target nrcogs to the degrees of such classifiers and chains.
2.4.2. Note that such explanations, as a side effect, could illuminate the limits of current means of cognizing for adequate representation of U* and, possibly, provide an additional, maybe more transparent axiomatic views on the fundamentals themselves.
2.5. Another way of approving the classifiers on the origination of nrcogs, i.e., the hypotheses on it, is an attempt to approve them constructively, namely, by constructing regularized classifiers Cl that adequately model the target ones Cl' as in [1].

2.5.1. *Regularized classifiers* Cl are accompanied by means, say, methods, procedures, and algorithms, that with some reliability not only identify the samples of Cl, but also reproduce, and generate such samples.
These means allow the production of realities equal to the samples of indoms of Cl either with some involvement of humans in their formation, or more without them, if automated.
2.5.2. Positives r of regularized classifiers Cl and Cl themselves are interpreted as *models* of classifiers Cl' if r are also classified as positives of Cl`, while Cl are interpreted as *adequate models* of Cl' if positives r meet certain additional requirements induced by positives of Cl'.
For example, if Cl classify algorithms, while Cl' computability, then, by Church, Cl adequately model Cl' if for any positive r' of Cl', equal positive r can be produced by Cl.
2.5.3. Classifiers Cl are *constructively regularized* if Cl are regularized and samples sps of Cl are assembled by explicit algorithms alCl from non-cellular independent units of matter.
And since algorithms alCl are capable of producing positives of Cl, they can, to some extent, equally substitute Cl.
2.5.4. The impact of constructively regularized classifiers Cl on the approval of hypothesis Cl' by its adequate modeling is based on the assumption that the provision by Cl with such modeling algorithms alCl acknowledges to some extent the revelation of cause-effect relationships comprising the nature of Cl'.
2.5.5. Note that the criteria of regularization comprise one of the cores of science. Particularly, we trust scientific classifiers Cl if for prescribed, reproducible conditions Cl procedurally provide certain predefined realities.
2.6. A mighty instrument of approval of classifiers Cl in science is the quantification of meanings of Cl.
The meanings of such quantified classifiers Cl include constituent classifiers, and *properties* accompanied by certain functions, operators, such as weighting, melting or boiling points, viscosity, density, etc., capable of corresponding to the positives of Cl in certain quantities. These quantities allow us to represent the positives of Cl quantitatively, i.e., to model Cl quantitatively, then, study these models in quantitative theories to interpret the results for the original Cl.
2.7. Thus, the approval of the origination of nrcogs could support explanations in proper theories, i.e., reliable chains of constructively regularized classifiers, as well as the construction of regularized classifiers adequately modeling nrcogs.
In practice, however, explanations along with regularized classifiers include ones experienced only with a variety of degrees of reliability up to the plain hypothesis.
Thus, in approving the origination of nrcogs in theories, at first, it could be available only to chain targets with adequately and constructively modeled classifiers, followed by attempts of approval of targets by massively experimenting and/or densely chaining them with already acknowledged reliable classifiers.
2.7.1. Keep in mind that despite the above ways of enhancing the reliability of approval of classifiers, eventually, classifiers cannot be absolutely reliable since they always extrapolate some restricted experiences.
2.7.2. Note also that, generally, if cause-effect chains are short-distanced from the fundamentals of theories, they explain *originations,* whereas, otherwise, they explain *formations.*

3. Takeaways from Originations

3.1. *Guiding takeaways*. Nowadays, the origination of realities tends to be interpreted by fundamentals of suitable theories.

It is acknowledged that quantum field theory (QFT) provides one of the most comprehensive views on origination and "…properties of nature at the scale of atoms and subatomic particles, complementing classical physics that describes many aspects of nature at an ordinary (macroscopic) scale but not sufficient for describing them at small (atomic and subatomic) scales" [8].

Reasonably, we can assume that QFT could be supportive in the interpretation of origination of non- cellular 1/2place classifiers and energizers, i.e., the nuclei of roots of cognizers (nrcogs), and will be looking for it as follows.

3.1.1. In theories such as QFT, assertions/classifiers are eventually based on experiences and assumptions, and thus vary in reliability as it was already introduced.

Postulates compared with other classifiers of theories, nevertheless, have certain preferences in reliability since they are massively experienced being only several allowed by cause-effect chains to infer an enormous amount of reliable classifiers of T, thus, ideally, capable to become necessary and sufficient for inferring the body of T.

Hence, targeting reliable explanations of the hypothesis on the origination of nrcogs, we need to position ourselves in postulates comprising the ground of reliability of suitable theories that, assumingly, can be referenced in explanations of our target nrcogs.

3.1.2. Note, that realities encompassing the already revealed primordial-based chains in theories, can appear as roots and/or constituents of chains of origination of nrcogs.

3.1.3. Note also that theories such as QFT signify one of the dimensions of physics and cannot, but are grounded on already accustomed classifiers of sciences and, moreover, on the entire human knowledge.

Thus, QFT interpretations and our takeaways from it will inevitably include common, accustomed units of community languages.

3.1.4. Keeping in mind the above notes, let us provide a takeaway from postulated and originated realities of QFT, as well as address the available premises of origination of nrcogs.

3.2. *Originations.* According to modern cosmology, the universe is expanding and there is convincing evidence that it was hot and dense in the past.

We can distinguish between the entire universe (U*), and the part (or the "patch") of the universe that we can observe. U* may contain many, perhaps infinite number of such patches, as there is no evidence of U* having any boundary. Thus, while any finite patch of U* would shrink to an infinitesimally small size in the past, U* could still contain an infinite number of such tiny regions and thus be infinite in extent [8,9].

When the density of the universe (both U* and any patch of it) becomes too high, quantum gravity effects are expected to dominate and the nature of space-time would change. Note, while there are some theoretical ideas on what may happen at that point, modern physics cannot yet describe the universe at such high densities.

3.2.1. All dynamical realities we see around us are, ultimately, made of elementary particles, such as electrons, quarks, and neutrinos that, in addition to gravity, interact via electromagnetic, weak and strong forces. Where did all these particles come from?

In the current understanding of particle physics, based on QFT, each type of particle corresponds to a fundamental quantum field that is postulated to have existed at all times.

What we perceive as a particle, is an excitation of the corresponding field. For example, an electron particle would be an excitation of the electron field.

Producing the excitations, i.e., the particles, requires energy. The energy could be transferred to the particle fields from other fields, such as the inflaton field responsible for driving a period of exponentially fast expansion in the early universe known as Inflation.

3.2.2. According to the inflationary paradigm, which is widely accepted by cosmologists, the universe experienced a period of rapid expansion in which the density of all particles was diluted to a negligible level.

This expansion was driven by the potential energy of the inflaton field – a fundamental field that is postulated to have certain properties that allow it to cause cosmic acceleration. The period of rapid expansion eventually ends when the inflaton starts to convert most of its potential energy into kinetic energy.
During this period, known as reheating, rapid oscillations of the inflaton field transfer energy to the particle fields, producing a large number of particles of all types that would be at very high temperature at that time.
Effectively, this is the moment of the Big Bang, when the universe became hot. At that time, all particles were massless.
The electrons and quarks do not acquire mass until later – this happens via the famous Higgs mechanism after the electroweak phase transition.
3.2.3. During Inflation, quantum fluctuations, that are inevitably present in all fields, are amplified by the rapid expansion and leave dents, or wells, in space-time after inflation ends. The wells serve as seeds for structures, such as stars and galaxies in the later universe.
As the universe expands, elementary particles assemble to form nuclei and atoms. The atoms then congregate in the wells left by Inflation and, through gravity, grow into larger clumps of matter that later form stars and galaxies.
3.3**.** *Evoking universes and energy.*
3.3.1. *Universes.* Let us recall the assumption that the universe U* is an extrapolation of those U of communities, in turn, comprised of universes based on experiences of particular observers, the members of communities that, eventually, are based on the imprints, i.e., the outputs of classifiers the members own at the time [1].
And *realities of members x of communities C* so far are defined as imprints of x, along with causers of imprints and their classifiers, while *the universe of the observers x*, xU, as totalities comprised of realities of x.

Uniting xU by members of communities C', we get *the universe of C', C'U*, and uniting C'U by all communities *the universe for all humans - HU, or U.*
And while U is regularized (not constructively) since the representation of U is regularly transferred through generations of humans, an extrapolated coverage U* of U, apparently, cannot be regularized.
3.3.2. *Energy* is postulated as a reality owned by any reality and, moreover, by anything [10].
Energy E(r) of realities r appears to observers as motion of r and/or their constituents, classified as *kinetic energy*. There is also energy that can *potentially be kinetic energy*, i.e., the *potential energy*, which, to become kinetic, needs to be released from its current reserved, restrained, bounded, mass and other appearances to observers.
3.3.3. Energy E(r) is quantifiable. For example, in classical mechanics, E(r) is measured by the work performed over the realities r1 to accelerate its mass from rest to its stated velocity and is expressed in joules or their equivalent derivatives.
3.3.4. Note that in QFT energy of realities r and their constituents can be measured if r explicitly are identified by systems s with totally nominated constituents.
Then, since the energy E(s) of systems s is invariant with respect to time translation symmetry, it should be measured for such representations of s that reflect this symmetry as it can do the wave function Ψ(s) of s.
Note that, what in classical physics used to be called physical quantity or measurable quantity, in QFT the standard term becomes *observable* to emphasize that the meaning of quantum realities must be specified by certain operators [11].
Examples of observables in quantum mechanics are position, velocity, momentum, angular momentum, spin, and energy.
Thus, at present, the Hamiltonian operator applied to Ψ(s) is used as a quantifier of the energy of systems s.

Namely, the measurement of energy E(s) of systems s follows the scheme:
r =>s => Ψ (s) => Input **[***Hamiltonian operator***]** Output =>E(s).

4. Premises to Origination of Dynamicity

4.1. QFT, being well tested and continuing to develop a theory, can provide reliable premises in interpreting target originations.
Indeed, QFT states that the realities r, charged with kinetic energy, for some reasons met with realities r1 causing a variety of changes of r1 and themselves changing, for example, the location of r1 in space, destroying or transforming r1 into other r1' ones.
Such acknowledgment allows us to reduce the question of origination of our targets to the question of what types of compounds can form, get-together such active, dynamic realities.
Following QFT primordial fields, being excited, originate particles, unite them in atoms, compounded in molecules that embrace the diversity of matter.
4.1.1. Note that the acknowledgment of existence of primordially dynamic realities in U* is the result of the reliable human experience, allowing us to question the origination of realities strongly in the frame of laws of science without any reference to extraterrestrial wills or intentions in U* associated with the existence of Divines or Gods.
4.2. Dynamic realities being one of the fundamentals of QFT are studied also in mechanics, chemistry, perception, processing of symbols, etc.
On the other hand, constructive models of cognizing – doers, are defined as a type of dynamic realities having input-output parts and for realities at the input parts elaborating certain output ones or remaining passive.
Correspondingly, 1/2place classifiers and energizers, interpreted as types of doers, are also dynamic realities charged with energy allowing them to process input realities into the output ones.
4.2.1. Thus, linking the dynamicity of doers with the dynamicity in sciences will allow us to enrich the dynamicity of 1/2place classifiers and energizers with those in sciences too.
Particularly, the goal is to enrich the doers and their modes by links with fundamentals such as energy attributed by forms of appearance, conservation and transition laws, measurement by work, and others, that could be supportive in their chaining with fundamentals, thus, helping to enlighten their origination.

5. Can Dynamicity Be Modeled by Doers?

5.1. The aim of linking the dynamicity of doers with those of fundamentals we refine, in general, questioning, whether the doers can be interpreted as models of dynamicity by QFT?
As premises and a footstep to a positive answer to this question in what follows, we provide cases of such interpretation of dynamicity in mechanics, chemistry, perception and processing of symbols by the dynamicity of doers, concluding with some hints to generalize the cases.
5.2. *Durables* in [1] were defined as realities that, in contrast with others, *temporalis*, can be properly identified in the meantime.
Durables are *stationaries* if the energy they possess in some forms or appearances is either stationer or its partial transition to some other forms or appearances can be ignored by observers.
Otherwise, durables are *dynamics*.
Stationaries, for example, are stones, rocks, pendulums or star systems.
Dynamics include rivers, oceans and, apparently, doers.
5.3. Parenting relationship in OOP can be interpreted as a statement.
St.1.5. *Classifiers Cl are parents for classifiers Cli, i=1,…,n, if all attributes of Cl are affirmative for all positives of Cli.*

Then, uniting all positives of Cli into those of classifier uCli, it can be stated as follows.
St.2.5. *Classifiers Cl parenting Cli, i=1,…,n are also parenting classifiers uCl of the unions of Cli.*
5.3.1. Classifiers Cli can be interpreted as sensors snCl, doins diCl [2], absorbers abCl, sugar synthesizers szCl and professionals prCl.
Indeed. A type of doers, classifiers, are parenting classifiers snCl of sensors that cause imprints from warmth, light, sound or chemical inputs, as well as the classifiers diCl of doins..
Classifiers abCl of some compounds of atoms absorbing certain chemicals, as well as those of szCl of synthesizing sugars from carbon dioxide and water, are also parented by doers as the type of classifiers.
Then, classifiers of doers are parents of classifiers prCl of human professionals, specialized in elaborating certain input realities into other ones. For example, loaders input some loads, cargos in some locations and relocate them, then, cookers input nutrients and output their processed modes, etc.
Correspondingly, we can state that
St.3.5. *Classifiers dCl of doers parenting classifiers snCl, diCl, atCl, abCl , szCl, prCl are also parenting their union udCl.*
5.3.2. Note, that classifiers snCl, diCl, atCl, abCl , szCl classify occasionally when they by chance get input positives, while professionals identified by classifiers prCl address some internal or external stimuli and become active intentionally.
5.4. Let us now address a dependency between parenting and the modeling of doer-classifiers.
Preliminarily, let us recall that
St.4.5. *If classifiers Cl are parenting classifiers Cli, i=1,…,n, and are regularized, then Cl become also the models of Cli and their union classifiers uCl.*
Indeed, classifiers dCl are regularized, either constructively or not, for classifiers snCl, diCl, atCl, abCl , szCl, prCl, udCl and are parenting them.
5.5. Note that atoms, in general, are only single and not recurrent classifiers, thus, they don't exactly meet the above statements. For example, atoms of hydrogen bound with some oxygen ones became unable to recurrently do the same for other oxygen atoms, as, seemingly, do the absorber abCl or compounds szCl synthesizing sugar seems, to represent such recurrent classifiers.

6. Conclusions

6.1. Relying on the above cases of positive interpretation of doers as models of dynamicity, it is worth questioning whether they can be models of dynamicity by QFT in general, i.e., be nuclei of dynamicity by QFT?
6.2. For this aim, it can be assumed that QFT doers r as a type of realities are charged by kinetic energy and reacting with some types of realities r1 they meet, i.e., realities r1 of their input domains, are transformed into realities (r1', r'), the output product of changing r1 and possibly r themselves. For example, r1' could be a new location of r1 in space, or be the result of destroying or transforming r1.
Then r' could be considered as renewed r after their interaction with r1.
6.2.3. For the gravity force field (gff) doers, it could be assumed that they are spread, distributed and acting at any point of the gravity field, as inputs could have any mass m at any position p, (m,p) and as output transpositions of (m,p) into (m,p1), where p1 could be attributed.
6.3. Other types of doers induced by QFT can be questioned, including those of interpreting particles as excitations of fields [11], as well as constituents of uncials [6].

References

[1] E. Pogossian, *Constructing Models of Being by Cognizing*, Academy of Sci. of Armenia, Yerevan, 2020.

[2] E. Pogossian, "Specifying adequate models of cognizers", American Institute of Physics , *AIP Conference Proceedings* 2757, 010001, doi:10.1063/12.0015299, 2023.

[3] Online.[Available]: https://en.wikipedia.org/wiki/Macromolecule

[4] M. Eigen, "Selforganization of matter and the evolution of biological macromolecules", *Naturwissenschaften*,vol. 58, pp. 465–523, 1977.

[5] B.-O. Küppers, *The Language of Living Matter. How Molecules Acquire Meaning*, Springer Nature Switzerland AG 2022.

[6] H. J. Peirce and communication, Kenneth L. Ketner (ed.) *Peirce and contemporary thought: Philosophical inquiries*, pp. 266, 1995.

[7] E. Pogossian, "Promoting origination of constituents of non cellular cognizers", *Proceedings of International Conference CSIT2023*, pp. 65-68, 2023.

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Fields*, Westview Press. 1995.

[9] E. Siegel, "Ask Ethan: How Do Quantum Fields Create Particles?", Online.[Available]: https://www.forbes.com/sites/startswithabang/2019/01/13/ask-ethan-how-do-quantum-fields-create-particles/?sh=719434b330ad

[10] Online.[Available]: https://en.wikipedia.org/wiki/Energy

[11] Online.[Available]: https://en.wikipedia.org/wiki/Field_(physics)

Ոչ բջջային իմացիչների դինամիկության առաջացման խթանում

Էդվարդ Մ. Պողոսյան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
e-mail: epogossi@aua.am

Ամփոփում

Դինամիկ իրողություններ գոյություն ունեն տարրական մասնիկների ամենաիմնական մակարդակում, որոնք, ըստ քվանտային դաշտային տեսության (quantum field theory), առաջանում են որպես հիմնարար քվանտային դաշտերի գրգռումներ: Միևնույն ժամանակ, իմացիչների (cognizers) միջուկները՝ գործիչները (doers) և նրանց եղանակները՝ 1/2 տեղանի դասակարգիչները և էներգիա առաքողները (energizers), նույնպես դինամիկ իրականությունների տեսակներ են: Մասնիկների և դաշտերի դինամիկայի մեջ գործիչների (doers) դինամիզմի ծագումը պարզելու փորձը կօգնի լույս սփռել բնության մեջ դասակարգիչների առաջացման վրա:Որպես այս հարցին դրական պատասխան տալու քայլ, մենք ներկայացնում ենք դինամիկության նման մեկնաբանության դեպքեր, որոնք ավարտվում են դեպքերն ընդհանրացնելու որոշ ակնարկներով:

Բանալի բառեր՝ Մոդելավորում, գործիչները (doers), էներգիայի առաքողներ (energizers), դասակարգիչներ, իմացիչներ (cognizers), հիմնարարներ, դինամիզմ:

Обосновывая возникновение динамизма неклеточных познавателей

Эдуард М. Погосян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армнения
e-mail: epogossi@aua.am

Аннотация

Динамические реалии существуют на самом базовом уровне элементарных частиц, которые, согласно квантовой теории поля (QFT), возникают как результат возбуждения фундаментальных квантовых полей. В то же время образующие неклеточных познавателей –акторы (doers) и их типы -1/2-классификаторы и энергизаторы, также являются типами динамических реальностей. Попытка проследить происхождение динамичности акторов (doers) в динамике частиц и полей может пролить свет на возникновение классификаторов в природе.

Обосновывая, что акторы (doers) могут быть моделями динамизма для примеров из ряда областей, мы задаемся вопросом, могут ли они быть моделями динамизма в более общем случае.

Ключевые слова: Моделирование, акторы (doers), энергизаторы, классификаторы, познание (cognizing), основания (fundamentals), динамичность.

Mathematical Problems of Computer Science 60, 27–34, 2023.
doi: 10.51408/1963-0106

UDC 510.5

Enhancing Symbolic Manipulation through Pairing Primitive Recursive String Functions and Interplay with Generalized Pairing PRSF

ShivKishan Dubey[1,2] and Narendra Kohli[2]

[1]Dr. Ambedkar Institute of Technology for Handicapped, Kanpur, 208024, Uttar Pradesh, India
[2]Harcourt Butler Technical University, Kanpur, 208002, Uttar Pradesh, India
e-mail: skd@aith.ac.in, nkohli@hbtu.ac.in

Abstract

In earlier studies, the notion of generalized primitive recursive string functions has been presented, and their connections with abstract paring-based primitive recursive string functions have been investigated. Our study is centered around establishing a fundamental theorem that states a connection between these two distinct sorts of functions. The theorem specifically establishes that the universal definition of every generalized pairing primitive recursive string function is contingent upon its correspondence with a conventional(abstract) pairing primitive recursive string function. This article introduces innovative concept of Pairing Primitive Recursive String Functions (P-PRSF) for manipulating and interacting with word pairs. Based on the principles of primitive recursion and pairing functions, P-PRSF enables the extraction, transformation, and combination of word components. The proposed theorems validate the effectiveness of P-PRSF in capturing relationships within word pairs. Moreover, the interplay between P-PRSF and Generalized Pairing PRSF (GP-PRSF) extends the concept to involve more intricate interactions.
Keywords: Word pairing, PRSF, Generalized Pairing PRSF, Superposition, Alphabetic PRSF.
Article info: Received 10 August 2023; sent for review 7 September 2023; received in revised form 3 November 2023; accepted 17 November 2023.

1. Introduction

The foundational principles of basic recursive string functions have provided a framework for comprehending the computing capacities of functions that manipulate individual words from a specified alphabet[1, 2, 3]. This novel idea introduces the notion of generalized pairing primitive recursive string functions (GP-PRSF) and endeavors to unveil the intricate connections between these advanced functions and their conventional counterparts, known as pairing primitive recursive string functions (P-PRSF) in the existing literature[4, 5].

The motivation behind this investigation stems from an inherent curiosity: what occurs when we surpass the limitations of individual words and engage in the process of manipulating pairs of words? In the context of this endeavor, the notion of GP-PRSF arises, enabling the integration of indeterminate functions inside this novel framework. This study aims to explore the fundamental connections between GP-PRSF and P-PRSF through an examination of a crucial theorem that provides insights into the circumstances under which these functions are uniformly defined.

2. Pairing Primitive Recursive String Function (P-PRSF)

The concept of Pairing Primitive Recursive String Functions (P-PRSF) emerges as a formalism to operate on pairs of words from a given alphabet. P-PRSFs build upon the foundation of traditional Primitive Recursive String Functions (PRSF) [5, 6] and extend their capabilities to address the complexities of word pairing.

Let $A = \{a_1, a_2, .a_p\}$ be a set of alphabets comprising $p > 1$ distinct symbols. The PRSF function F is defined as operating on pairs of words, (P, Q), where P and Q are words $\in A$. The aim of F is to produce an output based on the input pair. Further, introducing $\Pi_1(P, Q) = P$ and $\Pi_2(P, Q) = Q$ are basic pairing functions that extract the first and second components of the pair, respectively[5].

An undefined basic pairing function: In the context of the extension for word pairing as generalized word pairing (GPRPF), an undefined word pair function through $U(P, Q)$ returns an undefined value for any input pair (P, Q). It may refer to a scenario where a function is not defined for certain input pairs as similar to an undefined value for PRSF [7]. This could be due to possible specific conditions or restrictions on the input pairs and also serve as a foundation for generalized pairing.

2.1 Operators for Pairing Functions

Superposition: If F is a pairing function, and $G_1, G_2, , G_n$ are pairing functions, then the superposition F^* of F with $G_1, G_2, , G_n$ is defined[5] as Equation 1 below:

$$F^*(P, Q) = F(G_1(P, Q), G_2(P, Q), , G_n(P, Q)). \tag{1}$$

In the case of only two pairs of words $(P1, P2)$, it can be used as $F^*(P1, P2) = F(G_1(P1, P2), G_2(P1, P2))$, where only two representations are paired in terms of G_1 and G_2.

Alphabetic Primitive Recursion for Pairing Functions: If F is a pairing function, and $H_1, H_2, , H_p$ are $(n + 2)$ dimensional pairing functions, then the primitive recursive pairing function[5] F^+ of F with $H_1, H_2, , H_p$ is defined "for some" $1 \leq i \leq p$ as Equations 2 and 3:

$$F(P, Q) \qquad \text{if } R = \Lambda, \tag{2}$$

$$H_i(P, Q, R, F^+(P, Q, R)) \qquad \text{if } R = Qa_i \text{ for some } a \in A. \tag{3}$$

To better understand how operators use word pairing through P-PRSF, we perform these operations as according to Table 1, where we'll perform operations over three different input pairs. The output for each input pair can be understood as a word that has been constructed through a process that involves paired words and the application of functions. In the context

Table 1: Perform P-PRSF through superposition and Alphabetic PRPF over different input pairs.

Operations	**Definitions**	**Input1**	**Input2**	**Input3**
		$\langle ab, ba\rangle$	$\langle a, bb\rangle$	$\langle ba, aba\rangle$
Π_1		ab	a	ba
Π_2		ba	bb	aba
Superposition	with the following assumptions as $\begin{cases} G_1(P,Q) = \Pi_2 \\ G_2(P,Q) = S_g(Q) \\ \text{to extract the first symbol from the 2nd component of word pairs} \\ F(P,Q) = \Pi_1 \end{cases}$ $F^*(P,Q) : F(G_1(P,Q), G_2(P,Q))$	ba	bb	aba
Alphabetic PRPF	$F^+ = F(P,Q)$ as Equation 2 for $R = \Lambda$	ab	a	ba
Alphabetic PRPF	If $R = Qa_i$ as assumed **ab** $\begin{cases} H(P,Q,R,F^+(P,Q,R)), \\ F^+(P,Q,R) \text{ can be evaluated} \\ \text{recursively as} \\ \Pi_3(P1,P2,Qai) = \Pi_2(P1,P2)[7] \end{cases}$	H(ab, ba, **ab**, ba)	H(a, bb, **ab**, bb)	H(ba, aba, **ab**, aba)
Alphabetic PRPF	For simplicity(*), we assumed the H function to extract a word with repeated occurrences from input word pairs	ba	bb	aba

of word pairing and incorporating some assumptions in terms of G and H, this output simply represents an element that emerges from the relationships between the original input words, with the recursive process playing a role in shaping these outcomes.

Theorem 1. (P-PRSF for Word Pairing) *For any word pairing (P,Q) and given pairing functions F and G as defined, the Alphabetic Primitive Recursive String Function F^+ effectively captures and manipulates the relationships within word pairs.*

Proof. The previously proposed lemma [5] along with the proposed Theorem [5] was the foundation for the equivalence between GPRSF and PRSF. The connection lies in the concept of "S-image" or the superposition operation[4],[5],which transforms a string function into a new one. This concept is analogous to the idea of combining and manipulating word pairs using the proposed P-PRSF operations. Therefore, this theorem can be proved by induction on the length of the third component R of the input (P, Q, R).
Base Case: For $R = \Lambda$, the base case of Alphabetic Primitive Recursion applies $F^+(P, Q, \Lambda) = F(P, Q)$. This effectively extracts and manipulates the components of the input word pair according to F, demonstrating the foundational operation for word pairing.
Inductive Hypothesis: Assume that for any non-empty word $R = Qa$ of length n, $F^+(P, Q, Qa)$ effectively captures and manipulates the relationships within the word pair

(P, Q), guided by the pairing functions F and G. Consider the word $R = Qab$ of length $n + 1$. By the inductive hypothesis, $F^+(P, Q, Qab)$ is constructed through interactions between the word pair (P, Q), R, and the results of previous recursion steps.
Steps: i. Apply $H(P, Q, Qab, F^+(P, Q, Qab))$: This step involves interactions between P, Q, Qab, and $F^+(P, Q, Qab)$, effectively capturing relationships within the pair and guiding the construction of the result. ii. The recursive process navigates through the components of R and their interactions with F^+, eventually yielding a word that reflects the relationships between the original word pair.
By induction, for any non-empty word $R, F^+(P, Q, R)$ effectively captures and manipulates the relationships within the word pair (P, Q). The theorem is proven through induction, demonstrating that Alphabetic PRSF F^+ effectively captures and manipulate relationships within word pairs. ■

Theorem 2. (Preservation of Word Pairing Relations) *For any word pairing (P, Q) and given pairing functions F and G as defined, the Alphabetic PRSF F^+ preserves the inherent relationships and interactions within the word pair.*

Proof. $F^+(P, Q, R)$ is demonstrated to accurately preserve and reflect the relationships between the components of the input word pair (P, Q) and the recursive components R. Through Base Case: For $R = \Lambda$, $F^+(P, Q, \Lambda)$ evaluates to $F(P, Q)$, capturing the initial relationship between the components of (P, Q). Using Recursive Case: For $R = Qa$, where a is a symbol from the alphabet A, $F^+(P, Q, Qa)$ involves interactions between P, Q, Qa and the result of F^+ for the previous recursive component. This step accurately reflects the inherent interaction between the components of the word pair and guides the outcome based on the chosen pairing functions F and G.
Considering both the base and recursive cases, it becomes evident that $F^+(P, Q, R)$ effectively preserves the relationships and interactions within the word pair (P, Q) and the recursive components of R. The preservation of the word pairing relations is established through the accurate reflection of interactions and relationships within the word pair as guided by Alphabetic Primitive Recursive String Function F^+ . This mathematical proof reinforces the proposed idea of Pairing Primitive Recursive String Functions for word pairing scenarios. ■

Lemma 1. (Preservation of S-Image under P-PRSF Operations) *For any word pairing (P, Q) and given pairing functions F and G as defined, the application of Alphabetic PRSF F^+ to a pair of string functions F and G preserves the **S-image** property.*

Proof. Consider a string function F and its S-image F^*. Applying the Alphabetic PRSF F^+ on F and G to form $F^+(F, G)$, the new function $F^+(F, G)^*$ is obtained. As in Base Case: by applying F^+ to F and G retains the S-image property for $F^+(F, G)$, as the base case and operations of F^+ are defined consistently with the S-image property. Continue with Recursive Case: that preserves the S-image property under F^+, as it relies on the same underlying pairing functions F and G that maintain the S-image property.
By induction, the lemma demonstrates that the application of Alphabetic PRSF F^+ to the pairing functions F and G retains the S-image property, analogous to the preservation of S-image under generalized primitive recursive string functions. While the previous lemma (2015) deals with generalized primitive recursive string functions and their S-images, the

proposed new lemma focuses on the preservation of S-image under the proposed P-PRSF operations for word pairing. ■

Lemma 2. (Composition of P-PRSF is P-PRSF) *For any Pairing Primitive Recursive String Functions (P-PRSF) F and G, the composition $F \circ G$ is also a P-PRSF.*

Proof. The composition of P-PRSF functions F and G retains the properties of P-PRSF, namely the basic functions, superposition, and alphabetic primitive recursion. Both F and G are P-PRSF, they inherently preserve the basic functions (pairing functions Π_1 and Π_2) as well as alphabetic primitive recursion and superposition operations. Therefore, their composition $F \circ G$ also preserves the basic functions. The superposition operation is defined as $F^*(G^*(P,Q))$, where F^* and G^* are S-images of F and G, respectively. Both F^* and G^* are primitive recursive string functions in the usual sense due to the properties of P-PRSF. Since the composition of two primitive recursive functions is itself primitive recursive, $F \circ G$ retains the property of superposition. In respect of Alphabetic PRSF, F and G can also be defined as with respect to pairing functions Π_1 and Π_2. The composition $F \circ G$ is defined as $F^+(G^+(P,Q,R))$, where F^+ and G^+ are Alphabetic PRSF. The composition retains the property of alphabetic primitive recursion as it operates on the components and interactions of the word pair based on the P-PRSF operations.
By establishing the properties of basic functions, superposition, and alphabetic primitive recursion for the composition $F \circ G$, we conclude that the composition of P-PRSF functions is also a P-PRSF. This lemma demonstrates that the composition of P-PRSF functions adheres to the same principles and operations as individual P-PRSF functions. This supports the idea that the proposed operations for word pairing maintain their validity even when combined in a composite manner. ■

3. Interplay with Generalized Pairing PRSF (GP-PRSF)

Now extend the notion of Pairing Primitive Recursive String Functions (P-PRSF) to involve interactions with generalized versions of these functions[4],[6],[5]. This would allow us to combine the foundational idea of word pairing with more complex interactions based on GP-PRSF.
Let's denote the Generalized Pairing PRSF as $H(P,Q)$, where H is a function that involves interactions between word pairs (P,Q) and is guided by specific rules and operations. The interactions can be more intricate than basic pairing, incorporating additional considerations or conditions. Here, H is a function that takes two words P and Q as input and performs complex interactions between them according to H. The proposed concept involves applying Alphabetic PRSF F^+ to a pair of P-PRSF F and GP-PRSF H. This results in the formation of a new function $F^+(F,H)$, which captures the interplay between the simpler word pairing operations and the more intricate interactions guided by H. The focus is on demonstrating that the operations of $F^+(F,H)$ retain the desired properties of both P-PRSF and GP-PRSF, thus effectively combining the foundational and advanced aspects.

Example 1 *Now we'll define a GP-PRSF that measures the similarity between two words based on their character composition. It is actually a very well-developed idea where NLP researchers achieved effective results in this task, but our focus is only on the interplay between P-PRSF and GP-PRSF. We'll then apply the interplay concept to combine this*

GP-PRSF with a P-PRSF. Let's define a GP-PRSF $S(P,Q)$ that measures the similarity between two words P and Q based on their character composition. The function S calculates the number of common characters between P and Q, normalized by the length of the longer word. The formula for S can be defined[2] as the following equation:

$$S(P,Q) = \frac{Number\ of\ common\ characters}{max(length\ of\ P, length\ of\ Q)}.$$

Given a P-PRSF $F(P,Q)$ that extracts the first word of a pair ($F(P,Q) = P$), the interplay function $F^+(F,S)$ would involve applying $F(P,Q)$ to the first word of the pair and then calculating the similarity S between that word and the second word Q. Suppose we have the following word pair: P="apple" and Q="ample". Using The GP-PRSF S calculates the similarity between P and Q as 3/5=0.6, now to Interplay with GP-PRSF $F^+(F,S)$ P-PRSF $F(P,Q) = P$ to P, which results in P="apple". Then, we calculate the similarity S between P and Q as $F^+(F,S)(P,Q)$=$S(P,Q)$=0.6.
As above, The interplay between the basic pairing operation and the GP-PRSF involves extracting the first word P, calculating the similarity between P and Q, and obtaining a numeric value that represents the degree of similarity between the words.

4. Conclusion and Future Directions

The concept of Pairing Primitive Recursive String Functions (P-PRSF) within the domain of word pairing presents a systematic framework for the manipulation and interaction of word pairs. By integrating fundamental principles of basic recursion[1],[7] with the inventive methodology of pairing functions, this notion facilitates the extraction, modification, and integration of linguistic elements. The theorems and lemmas offered in this study provide evidence for the soundness of P-PRSF, demonstrating its efficacy in accurately representing and maintaining connections between word pairs. By examining the relationship between P-PRSF and the more detailed Generalized Pairing PRSF (GP-PRSF), the idea expands its practicality to encompass complex interactions. The notion of interplay provides opportunities to explore a wide range of applications, wherein fundamental word pairing procedures may be integrated with sophisticated interactions driven by particular rules. The inherent capacity of this interaction facilitates the development of novel functionalities that include the fundamental and intricate procedures.
The scope of future investigation may involve the advancement of algorithms that utilize the P-PRSF idea in order to facilitate activities such as text analysis, natural language processing, and data transformation. Potential applications of this technology include similarity calculations, text production, and pattern recognition. The study focuses on advanced interaction models that aim to investigate a range of GP-PRSF models for intricate interactions between pairs of words. This exploration has the potential to yield novel functionalities that effectively capture semantic links, contextually-aware transformations, and sentiment-based operations. The practical application of the P-PRSF idea to real-world challenges has the potential to yield innovative solutions. For instance, the use of P-PRSF can provide structured procedures[4],[5] that could be advantageous in the development of automated text editing tools, content summarizing techniques, and creative writing applications.

References

[1] A. I. Malcev, *Algorithms and Recursive Functions*, 2nd edition, Moscow, Nauka, (in Russian), 1986.

[2] I. D. Zaslavsky, "On some generalizations of the primitive recursive arithmetic", *Theoretical Computer Science*, vol. 322, pp. 221–230, 2004.

[3] V. Vuckovi, "Recursive word-functions over infinite alphabets", *Mathematical Logic Quarterly*, vol. 13, no.2, pp. 123-138, 1970.

[4] M. H.Khachatryan, "On the representation of arithmetical and string functions in formal languages'", *Transactions of the IIAP of NAS of RA, Mathematical problems of computer science*, vol. 27, pp. 37-53, 2006.

[5] M. H. Khachatryan, "On generalized primitive recursive string functions", *Mathematical Problems of Computer Science*, vol. 43, pp. 42-46, 2015.

[6] L. Santean, "A hierarchy of unary primitive recursive string-functions", *Proceedings 6th International Meeting of Young Computer Scientists, Aspects and Prospects of Theoretical Computer Science,* , Smolenice, Czechoslovakia, LNCS vol. 464, pp. 225-233, 1990.

[7] F. W.von Henke, G. Rose, K. Indermark and K. Weihrauch, "On primitive recursive word functions", *Computing*, vol. 15, no. 3, pp. 217-234, 1975.

Խորհրդանշական մանիպուլյացիայի բարելավում՝ զուգակցելով պարզունակ ռեկուրսիվ լարային ֆունկցիաները և փոխազդելով ընդհանրացված PRSF զուգավորման հետ

ՇիվԿիշան Դուբեյ[1] և Նարենդրա Կոհլի[2]

[1]Դոկտոր Ամբեդկարի տեխնոլոգիական ինստիտուտ հաշմանդամների համար, Կանպուր, Հնդկաստան
[2] Հարքուրթ Բաթլերի տեխնիկական համալսարան, Կանպուր, Հնդկաստան
e-mail: skd@aith.ac.in, nkohli@hbtu.ac.in

Ամփոփում

Ավելի վաղ ուսումնասիրություններում ներկայացվել է ընդհանրացված պարզունակ ռեկուրսիվ լարային ֆունկցիաների հասկացությունը, ինչպես նաև ուսումնասիրվել են դրանց կապերը վերացական զուգավորման վրա հիմնված պարզունակ ռեկուրսիվ լարային ֆունկցիաների հետ: Մեր ուսումնասիրությունը կենտրոնացած է հիմնարար թեորեմի հաստատման վրա, որը կապ է հաստատում այս երկու տարբեր տեսակի ֆունկցիաների միջև: Թեորեմը հատուկ հաստատում է, որ յուրաքանչյուր ընդհանրացված զուգավորման պարզունակ ռեկուրսիվ լարային ֆունկցիայի համընդհանուր սահմանումը պայմանավորված է նրա համապատասխանությամբ սովորական (վերացական) զուգակցվող պարզունակ ռեկուրսիվ տողային ֆունկցիայի հետ: Այս հոդվածը ներկայացնում է պարզունակ ռեկուրսիվ լարային ֆունկցիաների (P-PRSF) զուգակցման նորարարական հայեցակարգը՝ բառազույգերի հետ մանիպուլյացիայի և փոխազդեցության

համար: Հիմնվելով պարզունակ ռեկուրսիայի և զուգավորման ֆունկցիաների սկզբունքների վրա՝ P-PRSF-ը հնարավորություն է տալիս բառի բաղադրիչների քաղվածքը, փոխակերպումը և համակցումը: Առաջարկված թեորեմները հաստատում են P-PRSF-ի արդյունավետությունը բառազույգերի մեջ փոխհարաբերությունները ֆիքսելու հարցում: Ավելին, P-PRSF-ի և ընդհանրացված զուգակցման PRSF-ի (GP-PRSF) փոխազդեցությունը ընդլայնում է հայեցակարգը՝ ներառելով ավելի բարդ փոխազդեցություններ:

Բանալի բառեր՝ Բառերի զուգավորում, PRSF, ընդհանրացված PRSF զուգավորում, սուպերպոզիցիա, այբբենական PRSF:

Улучшение символьных манипуляций посредством объединения примитивных рекурсивных строковых функций и взаимодействия с обобщенным спариванием PRSF

ШивКишан Дубей[1] и Нарендра Кохли[2]

[1] Технологический институт для инвалидов доктора Амбедкара, Канпур, Индия
[2] Технический университет Харкорта Батлера, Канпур, Индия
e-mail: skd@aith.ac.in, nkohli@hbtu.ac.in

Аннотация

В более ранних исследованиях было представлено понятие обобщенных примитивно-рекурсивных строковых функций и исследованы их связи с абстрактными примитивно-рекурсивными строковыми функциями, основанными на спаривании. Наше исследование сосредоточено вокруг установления фундаментальной теоремы, которая устанавливает связь между этими двумя различными видами функций. Теорема конкретно устанавливает, что универсальное определение каждой обобщенной спаривающей примитивно-рекурсивной струнной функции зависит от ее соответствия обычной (абстрактной) спаривающей примитивно-рекурсивной струнной функции.

В этой статье представлена инновационная концепция объединения примитивных рекурсивных строковых функций (P-PRSF) для манипулирования парами слов и взаимодействия с ними. Основываясь на принципах примитивной рекурсии и функциях спаривания, P-PRSF позволяет извлекать, преобразовывать и комбинировать компоненты слова. Предложенные теоремы подтверждают эффективность P-PRSF при обнаружении отношений внутри пар слов. Более того, взаимодействие между P-PRSF и Generalized Pairing PRSF (GP-PRSF) расширяет концепцию и включает более сложные взаимодействия.

Ключевые слова: Пары слов, PRSF, обобщенные пары PRSF, обобщенные пары PRSF, суперпозиция, алфавитные PRSF.

Mathematical Problems of Computer Science 60, 35–39, 2023.
doi: 10.51408/1963-0107

UDC 519.2

On Testing of Multiple Hypotheses of Continuous Probability Distributions Arranged into Two Groups

Aram O. Yesayan

Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
French University in Armenia (UFAR), Yerevan, Armenia
e-mail: aram.yesayan@pers.ufar.am

Abstract

The optimal Neyman-Pearson procedure of detection is investigated for models characterized by four continuous probability distributions arranged into two groups considered as hypotheses. It is worthy to note that the case of three discrete probability distributions arranged in two groups was studied by Haroutunian and Yesayan in [1]. The Neyman-Pearson theorem holds immense importance when it comes to solving problems that demand decision making or conclusions to a higher accuracy.
Keywords: Neyman-Pearson procedure, Continuous probability distribution, Probability density function, Hypothesis testing, Error probabilities, Reliabilities.
Article info: Received 10 March 2023; sent for review 16 March 2023; received in revised form 8 October 2023; accepted 13 November 2023.

1. Introduction

The Neyman-Pearson theorem states that the likelihood ratio test is the most powerful test for a given significance level (or size) in the context of simple binary hypothesis testing (null hypothesis against alternative hypothesis) problems. It provides a theoretical basis for determining the critical region or decision rule that maximizes the probability of correctly detecting a true effect while maintaining a fixed level of Type I error.

Statistical power represents the ability of a hypothesis test to detect a true effect or difference when it exists in the population. The theorem emphasizes the importance of optimizing this power while controlling the risk of both Type I and Type II errors. Type I error, also known as a false positive, occurs when we reject the null hypothesis (assuming an effect or difference exists) when it is actually true. Type II error, on the other hand, refers to a false negative, where we fail to reject the null hypothesis (assuming no effect or difference) when an effect or difference truly exists. The Neyman-Pearson theorem allows us to strike a balance between these errors by maximizing power while setting a predetermined significance level (the probability of Type I error).

In [2]-[4], Cox formulated several divers examples of problems for two families of hypotheses testing and developed a general modification of the Neyman-Pearson maximum-likelihood

ratio procedures for the solution of such problems for the parameters of known continuous probability distributions (CPDs). In [1], Haroutunian and Yesayan studied the problems concerning the Neyman-Pearson criterion where discrete probability distributions are arranged in many groups and where the error probabilities decrease exponentially as 2^{-NE}, when the number of observations N (size of sample) tends to infinity. In [5], Tusnády studied the hypotheses testing problem of two CPDs, where error probabilities also exponentially approach zero. The optimal hypotheses testing problems, when error probabilities exponentially approach zero were also studied in [6] and in [7]-[9]. In [8], Haroutunian, Hakobyan and Hormosi-nejad studied on two-stage optimal testing of multiple hypotheses for the pair of families of discrete distributions. In [9], Yesayan and Gevorgyan solved the problem of many CPDs by means of two-stage asymptotically optimal testing of multiple hypotheses based on Tusnády's result.

The hypotheses testing problems for two hypotheses were described in detail by Borovkov [10], Levy [11], van Trees [12], Csiszár and Longo [13], Csiszár and Shields [14], Longo and Sgarro [15]. The Neyman-Pearson criterion of multiple hypotheses testing for discrete random variable was explored in [16].

This paper is devoted to the generalization of the Neyman-Pearson criterion for composite hypotheses testing problem of CPDs. The result is based on the method proposed by Thomas and Cover [17] in the paragraph of information theory and statistics.

2. Problem Presentation and Solution

Let $\mathcal{P}(\mathcal{X})$ be the space of all CPDs. Let X be a continuous random variable (CRV) with one of 4 possible CPDs given by probability density functions (PDFs) $f_m,\ m = \overline{1,4}$. Let $\mathbf{x} = (x_1, x_2, ..., x_N)$, $x_n \in \mathcal{X},\ n = \overline{1,N}$, be a vector of results of N independent observations of the RV X, then the PDF will be $f_m^N(\mathbf{x}) = \prod_{n=1}^{N} f_m(x_n)$.

For a CRV X, four PDFs $f_1,\ f_2,\ f_3,\ f_4$ are given, called the hypotheses. A statistical hypothesis H is a conjecture about the distribution of population X.

The statistician should make a decision about CPD of CRV. In this paper, we consider this problem in two stage. These PDFs are divided into two groups (hypotheses) such that the first hypothesis H_1 is the group of $k = 1, 2, 3$ PDFs and the second hypothesis is the group of $4 - k$ PDFs. Let us consider the partition when $k = 2$ and the hypotheses are as follows:

$$H_1 : \{f_1,\ f_2\},\ \ H_2 : \{f_3,\ f_4\}. \tag{1}$$

In the first stage the statistician must accept or reject the first hypothesis on the base of sample $\mathbf{x}$. If the first hypothesis is not rejected the statistician can detect which PDF (f_1 or f_2) corresponds to CRV. So, if it is rejected the second stage detection will be between f_3 and f_4.

Taking decisions about the hypotheses statistician can commit the errors.

The probability $\alpha_{l|l}^N$ is to accept a hypothesis different from the true hypothesis H_l, $l = 1, 2$.

We will show that the proposed Thomas and Cover's proof of Neyman-Pearson theorem for discrete probability distributions will also work for this case.

We will use these notations for the *maximum-likelihood* ratio procedure, so we will take the *maximum* of a pair of PDFs: $g_1^N(\mathbf{x}) = \max({f_1}^N(\mathbf{x}); {f_2}^N(\mathbf{x}))$, $g_2^N(\mathbf{x}) = \max({f_3}^N(\mathbf{x}); {f_4}^N(\mathbf{x}))$.

Theorem 1. *For the threshold $t \geq 0$, consider the test Ψ_N^* defined by region of acceptance $\mathcal{A}^{N*}$ for hypothesis H_1:*

$$\mathcal{A}^{N*} = \left\{ \mathbf{x} : \frac{g_1^N(\mathbf{x})}{{g_2}^N(\mathbf{x})} > t \right\},$$

and acceptance region $\overline{\mathcal{A}^{N*}}$ for H_2.

So, by these definitions, the corresponding error probabilities (mentioned also in the introduction) will be

$$\alpha_{1|1}^{N*}(t) = \alpha_{2|1}^{N*}(t) = g_1^N(\overline{\mathcal{A}^{N*}}) = \int_{\mathbf{x} \in \overline{\mathcal{A}^{N*}}} g_1^N(\mathbf{x}) d(\mathbf{x}),$$

$$\alpha_{2|2}^{N*}(t) = \alpha_{1|2}^{N*}(t) = g_2^N(\mathcal{A}^{N*}) = \int_{\mathbf{x} \in \mathcal{A}^{N*}} g_2^N(\mathbf{x}) d(\mathbf{x}).$$

Let $\mathcal{A}^N \subset \mathcal{X}^N$ be the decision region for H_1 of another test Ψ_N with error probabilities $\alpha_{1|1}^N$ and $\alpha_{2|2}^N$. If $\alpha_{1|1}^N \leq \alpha_{1|1}^{N*}$, then $\alpha_{2|2}^N \geq \alpha_{2|2}^{N*}$.

Proof. Let $\Psi_{\mathcal{A}^{N*}}$ and $\Psi_{\mathcal{A}^N}$ be indicator functions of regions. The indicator function is 1, if the sample belongs to the corresponding region, and 0, otherwise. It is obvious that for all $\mathbf{x} \in \mathcal{X}^N$,

$$(\Psi_{\mathcal{A}^{N*}}(\mathbf{x}) - \Psi_{\mathcal{A}^N}(\mathbf{x}))(g_1^N(\mathbf{x}) - t{g_2}^N(\mathbf{x})) \geq 0.$$

Then

$$\int_{\mathbf{x} \in \mathcal{X}^N} (\Psi_{\mathcal{A}^{N*}}(\mathbf{x}) g_1^N(\mathbf{x}) - t\Psi_{\mathcal{A}^{N*}}(\mathbf{x}) {g_2}^N(\mathbf{x}) - \Psi_{\mathcal{A}^N}(\mathbf{x}) g_1^N(\mathbf{x}) + t\Psi_{\mathcal{A}^N}(\mathbf{x}) {g_2}^N(\mathbf{x})) d(\mathbf{x})$$

$$= \int_{\mathbf{x} \in \mathcal{A}^{N*}} (g_1^N(\mathbf{x}) - t g_2^N(\mathbf{x})) d(\mathbf{x}) - \int_{\mathbf{x} \in \mathcal{A}^N} (g_1^N(\mathbf{x}) - t g_2^N(\mathbf{x})) d(\mathbf{x})$$

$$= (1 - \alpha_{1|1}^*) - t\alpha_{2|2}^* - (1 - \alpha_{1|1}) + t\alpha_{2|2} = (\alpha_{1|1} - \alpha_{1|1}^*) + t(\alpha_{2|2} - \alpha_{2|2}^*) \geq 0.$$

So, from $\alpha_{1|1} \leq \alpha_{1|1}^*$ it follows that $\alpha_{2|2} \geq \alpha_{2|2}^*$.

3. Conclusion

This paper discussed a suitable strategy of hypotheses testing for models with 4 known CPDs grouped in 2 clusters, considered as hypotheses. This problem can be generalized for $M > 4$ hypotheses, which can be grouped into 2 clusters in various combinations, i.e., the first hypothesis will be composed by $K = 1, 2, ...M - 1$ PDFs and the second by $M - K$ PDFs. The solving method will be the same, but it is obvious that the result of each combination will be different.

References

[1] E. A. Haroutunian and A. O. Yesayan, "A Neyman-Pearson proper way to universal testing of multiple hypotheses formed by groups of distributions", *Mathematical Problems of Computer Sciences*, vol. 54, pp. 18-33, 2020.

[2] D. R. Cox, "Tests of separate families of hypotheses" *In Proceeding 4th Berkley Simp. Math. Statist. Prob., University of Callifornia Press, Berkelly*, pp. 105-123, 1961.

[3] D. R. Cox, "Further results on tests of separate families of hypotheses", *Journal of the Royal Statist. Society: Serie B*, vol. 24, issue 2, pp. 406-424, 1962.

[4] D. R. Cox, "A return to an old paper: Tests of separate families of hypotheses", *Journal of the Royal Statist. Society: Serie B*, vol. 75, issue 2, pp. 207-2015, 2013.

[5] G. Tusnády, "On asymptotically optimal tests", *Annals of Statatistics*, vol. 5, no. 2, pp. 385-393, 1977.

[6] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, vol. 36, pp. 369–401, 1965.

[7] E. A. Haroutunian, "Reliability in multiple hypotheses testing and identification problems" *Nato Science Series III, Computer and System Sciences*, vol.198, IOS Press, pp. 189-201, 2003.

[8] E. A. Haroutunian, P. M. Hakobyan and F. Hormosi-nejad, "On two-stage LAO testing of multiple hypotheses for the pair of families of distributions", *Journal of Statistics and Econometrics Methods* , vol. 2, no. 2, pp. 127-156, 2013.

[9] A. O. Yesayan and H. Z. Gevorgyan,On two-stage testing of multiple hypotheses testing concerning continuous random variable , *Europian academia: Collection of scientific articles*, vol. 7, pp. 294-299, 2016.

[10] A. A. Borovkov, *Mathematical Statistics (in Russian)*. Nauka, Novosibirsk, 1997.

[11] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, Springer, 2008.

[12] H. van Trees, *Detection, Estimation and Modulation Theory*, pt.1 New York, Wiley, 1968.

[13] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses", *Studia Sc. Math. Hungarica*, vol. 6, pp. 181–191, 1971.

[14] I. Csiszár and P. Shields, " Information theory and statistics: A tutorial", *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, 2004.

[15] G. Longo and A. Sgarro, " The error exponent for the testing of simple statistical hypotheses: A combinatorial approach", *Journal of Combinatorics and Informational System Science*, vol. 5, no. 1, pp. 58-67, 1980.

[16] E. A. Haroutunian and P. M. Hakobyan, "On Neyman-Pearson principle in multiple hypotheses testing", *Mathematical Problems of Computer Sciences*, vol. XL, pp. 34-37, 2013.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, New York, Wiley, 2006.

Երկու խմբում դասավորված անընդհատ հավանականային բաշխումների վերաբերյալ բազմակի վարկածների ստուգում

Արամ Օ. Եսայան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
Հայաստանում Ֆրանսիական համալսարան, Երևան, Հայաստան
e-mail: aram.yesayan@pers.ufar.am

Ամփոփում

Նեյման-Պիրսոնի ստուգման օպտիմալ ընթացակարգը հետազոտվում է այն մոդելների համար, որոնք բնութագրվում են չորս անընդհատ հավանականային բաշխումներով, որոնք բաժանված են որպես վարկածներ դիտակվող երկու խմբի մեջ: Հարկ է նշել, որ երկու խմբի մեջ բաժանված երեք դիսկրետ հավանականային բաշխումների դեպքը հետազոտվել է Հարությունյանի և Եսայանի կողմից [1]:

Նեյման-Պիրսոնի թեորեմը մեծ նշանակություն ունի, երբ խոսքը վերաբերում է այնպիսի խնդիրների լուծմանը, որոնք պահանջում են որոշումներ կայացնել կամ ավելի բարձր ճշգրտությամբ եզրակացություններ անել:

Բանալի բառեր՝ Նեյման-Պիրսոնի տեստ, անընդհատ բաշխում, խտության ֆունկցիա,վարկածների ստուգում, սխալի հավանականություն, հուսալիություն:

О проверке многих гипотез непрерывных распределений вероятностей расположенных в двух группах

Арам О. Есаян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
Французский университет в Армении, Ереван, Армения
e-mail: aram.yesayan@pers.ufar.am

Аннотация

Исследуется оптимальная процедура тестирования Неймана-Пирсона для моделей, характеризующихся четырьмя непрерывными распределениями вероятностей, разбитыми на две группы, рассматриваемые как гипотезы. Примечательно, что случай трех дискретных распределений вероятностей, расположенных в двух группах, был изучен Арутюняном и Есаяном [1].

Теорема Неймана-Пирсона имеет огромное значение, когда речь идет о решении задач, требующих принятия решений или выводов с более высокой точностью.

Ключевые слова: Критерий Неймана-Пирсона, непрерывное распределение, функция плотности, проверка гипотез, вероятность ошибки, надежность.

Mathematical Problems of Computer Science 60, 40–51, 2023.
doi: 10.51408/1963-0108

UDC 004.8

Multi-Stage Classification Scheme to Optimize Medical Treatments

Karen M. Gishyan

Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
e-mail: karengishyan.res@outlook.com

Abstract

In most existing machine learning and deep learning settings, classification and regression prediction problems may be described as a process where the model output is based on a single-stage input. In most real-life scenarios achieving the desired medical state for the patient may involve dynamically solving drug prescription problems based on the input data at different stages, where each stage is a logical grouping such as timestep division, ICU stay, etc. Data at a given stage represents a recovery progression and can be fundamentally different from the datasets from the previous and future stages. Although A single model may solve the task, a multi-stage learning procedure may be more suitable. To solve this task, we propose an FNN-driven ensemble-based approach for predicting the medications that the patient should receive at each stage of the recovery process. The final medical discharge location is predicted as a result of sequential predictions of drugs and features. In this work, we combine model ensembling and multi-stage iterative learning for solving an optimal drug prescription generation task as a contribution to the existing literature.
Keywords: Multi-stage classification, Treatment-optimization, Model-ensembling, Machine learning.
Article info: Received 21 March 2023; sent for review 2 May 2023; received in revised form 3 July 2023; accepted 4 September 2023.
Acknowledgement: The author would like to thank some of the department members for their valuable feedback.

1. Introduction

In this work, we propose a feedforward neural network-based iterative ensemble model architecture for solving the target class classification task for the medical domain, where the target class is the *Home Discharge* [1]. The proposed approach involves *Feature*, *Drug*, and *Output* prediction networks, which can be considered as an extension of decomposition-based (e.g., divide and conquer) and multi-objective optimization-based ensemble methods. Similar to

[1]Code source: `https://github.com/karen-gishyan/project`

some of decomposition methods, the original dataset is divided into a collection of datasets for multiple sub-processing, and similar to multi-objective optimization, we output multiple diverse predictors instead of a single predictor. Unlike decomposition methods, single *Drug* and *Feature* networks are optimized on different datasets at different stages instead of multiple predictors being optimized of the same type. This approach provides flexibility in adding more stages to the model architecture. This also means that the same set of network parameters is updated multiple times during a single forward pass. The alternative could be to construct a different network for each of the input datasets for each stage, where the downside would be that the number of networks that need to be optimized linearly would increase with each additional stage in the model architecture. Training is divided into three fixed stages (steps). For each stage s except the last one, we train two distinct FNN models: the *Drug* prediction network and the *Feature* prediction network, which predict drugs and features for the next stage $s + 1$, respectively. The feature and drug prediction training inputs at stage $s + 1$ are concatenated with the features and drugs at stage s. The rationale behind this approach is that the drugs given to a patient at each stage depend not only on the drugs at the previous stage but also on the patient's features at the previous stage. The same logic applies to the features, which depend not only on the past features but also on the past medication. We believe this approach results in more diverse datasets and better learning. Currently, the actual datasets for model training are available only at the first stage, and the training datasets at each stage s are iteratively generated based on the predicted features and drugs using only the observed data at the first stage as an initialization point. For the last stage, we train an *Output* model for learning the discharge location classes. Overall, we use three models across all stages: *Drug* network, *Feature* Network, and *Output* network. The main method can be considered a model-based approach with a strong emphasis on data-driven stage-based logic, data components, and logical variations of which were first described in [1] and [2] and further detailed in this work. The contributions can be summarized in the following points:

- We provide an ensemble-based iterative classification/regression pipeline that includes not just one but three different networks, each being optimized simultaneously in the forward pass. In addition, each *Feature* and *Drug* network is used multiple times depending on the number of stages. This is different from some of the existing approaches, where a new network is initialized at each stage.

- Having patient features for time t, we predict treatment for $t + n$ periods, where n is the number of stages (3 for this experiment).

- We train our models on synthetic datasets and give a detailed description of the stage-based data preprocessing technique.

2. Related Work

Ensemble methods have been widely used in research fields such as computational intelligence and machine learning. Ensemble methods can be categorized into conventional ensemble methods such as bagging, boosting and random forest, decomposition, negative correlation learning multi-objective optimization-based methods, fuzzy ensemble, multiple kernel learning ensemble, and deep learning ensemble Diversity is important in ensemble methods, and the three ways to create diversity are data diversity, parameter diversity,

and structural diversity [3, 4, 5]. Data diversity creates multiple datasets from the input dataset to train different models. The more diverse the datasets, the more diverse the model learning. Parameter diversity uses different parameter sets for generating different base predictors, and even with the same training set, the output of the predictors may differ. In structural diversity, ensemble predictors have different structures and architectures, and this kind of ensemble is also known as a heterogeneous ensemble [3]. Besides data, parameter, and structural diversity methods, there are other methods such as divide and conquer [6], multi-objective optimization [7], and fuzzy ensemble. In multi-objective classification, the training process yields a collection of optimal and diverse predictors instead of a single predictor [3]. Divide and conquer is mostly seen in time series forecasting, where the original dataset is often divided into a collection of parallel or hierarchical datasets, forming subtasks. Predictors are applied to each subtask, and then the outputs are aggregated. Datasets usually have different characteristics, and predictors mainly differ from each other. In divide and conquer methods, the original time series is decomposed into a collection of time series from which the original series can be reconstructed [6]. The goal is to obtain smaller and simpler time series, apply predictions to the decomposed time series, and aggregate the predictions. Both seasonal decomposition and wavelet transform are decomposition-based ensemble methods. While the first approach implies a prediction algorithm such as an SVR being applied to each seasonally decomposed component, in the second approach, a prediction algorithm is applied to the sub-series obtained by decomposing the original series into orthonormal series by the time domain [6]. [8] presents a divide-and-conquer-based hierarchical optimization framework for ensemble classifier learning. The framework includes a data training environment (DTE) creation that divides the data into multiple clusters and then trains heterogeneous base classifiers, which are later combined for an optimal ensemble. For optimizing multi-stage cascade classifiers, [9] proposes a deep model, which jointly optimizes multiple classifiers through several stages of backpropagation. Cascade classifiers were first proposed in [10] for solving a multi-stage recognition problem. Since then, cascading classifiers have been successfully applied to tasks such as image recognition [11], name entity recognition in clinical notes [12], anomaly detection and localization [13], and so on.

There are a few examples of multi-stage classification used in the medical domain. [14] solves a multi- stage classification problem for HER2 breast cancer by proposing a transfer learning-based approach used on the BCI dataset. [15] proposes an effective feature ensemble with multi-stage classification for breast cancer diagnosis, and the verification happens on a publicly available mammogram image dataset collected from the IRMA project. [16] proposes an automatic system involving multi-stage classification for diagnosing congestive heart failure using short-term heart rate variability analysis. For the experiments, open databases from Physionet, Normal Sinus Rhythm Database (NSR2DB), and Congestive Heart Failure Rhythm Database (CHF2DB) are used. [17] uses a multi-stage approach for performing arrhythmia recognition and classification. [18] uses a machine learning-based multi-stage classification method to classify Alzheimer's disease more efficiently. [19] uses a two-stage machine learning classification approach for heart disease prediction. [20] proposes a two-stage multi-modal learning algorithm for multi-label skin disease classification. [21] proposes a multi-stage approach to detect tumors, classify them into glioma or meningioma and perform their segmentation. [22] uses a transformer-based model for automatic multi-stage classification of diabetic retinopathy. [23] uses multi-stage superpixel classification for classifying four lung diseases and healthy lungs using chest X-ray images.

3. Data

For the experiments from the *MIMIC-III* clinical database, we use *ADMISSIONS*, *D_Items*, *PRESCRIPTIONS*, datasets, and a subset of the *CHARTEVETS* dataset, the latter containing 5 million rows from the whole dataset. We use the datasets listed above for generating *Features* and *Drugs* datasets. The *Features* dataset includes admissions and a list of those features for which the patient has had between 10 and 300 measurements throughout the stay. Similar logic is applied to the *Drugs* dataset, where we select the admission for which the patient was given more than 10 drugs. These two datasets are further filtered based on the admission IDs that are present in both. This approach is performed for the *DIABETIC KETOACIDOSIS* diagnosis.

3.1 Stage-Based Features and Drugs

As a result of data processing, where we generate the initial versions of the *Features* and *Drugs* datasets, the results of which can be observed in Table 1, we proceed to do extra stage-based processing to obtain the datasets for each time-stage.

3.1.1 Step 1 Processing

From *Features* and *Drugs*, we filter those observations where the patient's stay length was between 6 and 8 days. We define three stages and generate one *Features* dataset, and one *Drugs* dataset for each stage and one *Output* dataset only for the third stage. The first stage is defined as the *Initial* stage, the second as the *Intermediary* stage, and the third stage as the *Final* stage. The features are *O2 saturation pulse oximetry*, *Heart Rate*, *Respiratory Rate*, *Non-Invasive Blood Pressure mean*, *Non-Invasive Blood Pressure systolic*, *Non-Invasive Blood Pressure diastolic*, *Temperature Fahrenheit*, *Arterial Blood Pressure systolic*, *Arterial Blood Pressure diastolic*, *Arterial Blood Pressure mean*. We define a *stage* to be a period corresponding to 2 days spent at the hospital for the first two stages and between 2 and 4 days spent at the hospital for the final third stage. This means that the features observed and drugs given for the first 2 days become the *Features* and *Drugs* datasets for the first stage, the ones for the 3rd and 4th days the datasets for the second stage, and from the 5th day up to the 6th or to the 8th day, depending on admission, the datasets for the third stage. The rationale behind choosing these numbers for defining the stages is that we want each stage to have at least 2 days' worth of data. The final stage can have up to two days of more data compared to previous stages to loosen the restriction for the total admission duration to be precisely 6 days. However, a bigger difference in the number of days between stages means more time series observations for some stages compared to others, and we keep it to 2 to avoid increasing the tradeoff further. The more the difference between observations, the more value padding should be performed to make sure that the datasets have the same shape.

At this point, we have 3-dimensional data, where the first dimension (batch) is the given admission, the second dimension (rows) is the number of time steps for each of the given stage, and the third dimension (columns) is the features or the drugs, depending on if the dataset is the *Features* or the *Drugs*. The structure of the *Features* dataset can be seen in Table 2.

Table 1. Initial preprocessing results.

Dataset Statistics		
	Features Dataset	**Drugs Dataset**
Original		
N Observations	2949897	4945985
Unique Admissions	3869	47031
Unique Diagnoses	1445	13880
After Filtering		
N Observations	2948538	474213
Unique Admissions	3859	3859
Unique Diagnoses	1443	1443

Table 2. *Features* dataset for a single batch before averaging.

Features Dataset				
Timestep	**Feature 1**	**Feature 2**	$\cdots$	**Feature 10**
t_1	value 1	value 1	$\cdots$	value 1
t_2	value 2	value 2	$\cdots$	value 2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_{10}	value 10	value 10	$\cdots$	value 10

3.1.2 Step 2 Processing

Processing of this stage allows us to obtain datasets that will be used as a basis of synthetic data generation used for modeling and experimentation. The processing described in Section (3.1) has one limitation. Each patient will surely have a different number of drugs given and a different number of charted feature measurements for each stage. As most deep learning frameworks, including the one used in this work, assume that each batch input (admission data) for the model has the same shape, this means that all the batches need to be padded with a predefined value for the input data to have a certain shape of (i, j, k). After padding, we remove the second dimension by averaging the time series instances over the rows, which we believe makes the data for each stage more representative and less dependent on a single time-stage observation, which in most cases would be an artificially padded value. For the *Drugs* dataset, each drug is first encoded and is assigned a discrete numerical value, but as there is no natural ordering between the drugs, we further do a dummy conversion, which means each drug will be present for each patient in the form of either 0 or 1 (absent or present) showing whether the given drug was part of the patient's treatment procedure for the given stage. At this stage, we also generate an *Output* dataset, which again holds values of 0 or 1, which are later used for the binary classification task. To summarize, we end up with three datasets. In stages 1, 2, and 3, we get *Features*, *Drugs*, and *Output* datasets with

the following shapes $(n, 10)$, $(n, 902)$, and $(n, 1)$, respectively, where n is the number of rows or unique admissions. Note that the *Output* dataset is only present for the 3rd stage.

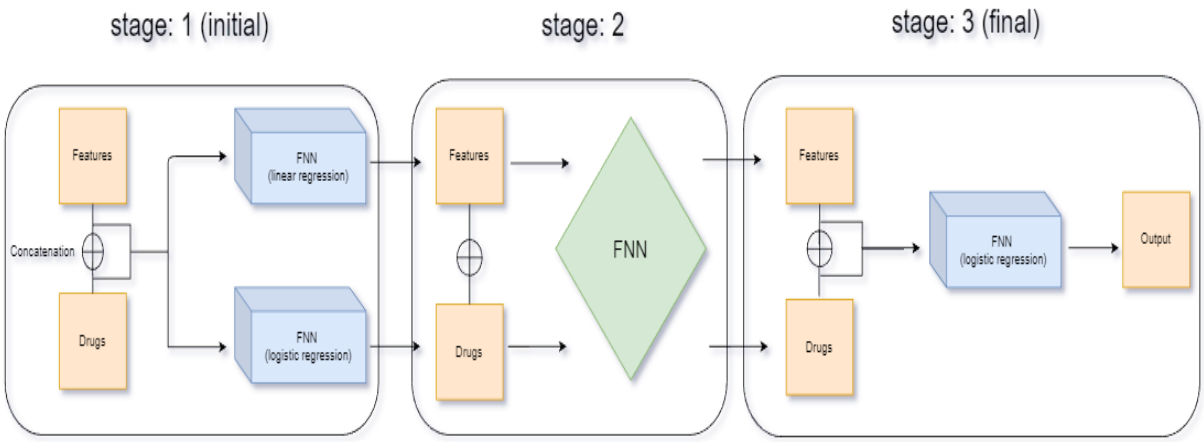

Fig. 1. Multi-stage classification pipeline.

4. Methodology

4.1 Synthetic Data Generation

Due to the particular data processing logic, the *Drugs* and *Features* datasets that we obtained for *DIABETIC KETOACIDOSIS* have only 16 observations, a shape of $(16, 902)$. To be able to train a machine learning model, we augment our datasets using a synthetic data generation technique using the Synthetic Data Vault (SDV) package [24]. We use a TVAE model, a VAE-based deep learning data synthesizer [25]. We augment each dataset to a size of 5000 observations; however, not all the observations are eventually used for model fitting. The average synthesized *Features* dataset similarity to the original *Features* dataset across three stages is 64%, while the synthesized *Drugs* dataset similarity to the original *Drugs* dataset across three stages is 98%. Output is generated as part of the *Features* for the third stage. As you can see, the similarity is very high for the *Drugs* dataset, which is understandable as the data is in a binary format compared to the continuous *Features* dataset. There is still, however, a significant class imbalance in the augmented datasets. For demonstration, there are only 184 instances where the final discharge outcome was positive (labeled as 1), while for the rest of the observations, the labeling is 0. The model training suffers from such imbalance, and the training results are poor. To overcome this problem, from dataset instances that have discharge outcomes of 0, we filter the first 184 instances and concatenate them to the other 184 instances that have positive discharge outcomes, and we obtain perfectly balanced datasets. This is performed for both *Drugs* and *Features* for each timestep. One may observe that the augmented datasets are significantly reduced, but we do this action willingly to make sure that the model results are attributable to the model itself and do not suffer from poor data. We still do acknowledge that results from synthetic datasets may not be fully representative of the original datasets.

4.2 Networks and Evaluation

To solve the problem of multi-stage classification, we construct an Ensemble pipeline comprising three networks; *Feature* prediction network, *Drug* prediction network, and *Output*

prediction network. Each network is a Feedforward Neural Network (FNN) model. The first network predicts patient features at the next stage, while the second network predicts patient drugs at the next stage. Each network takes a concatenated matrix of features and drugs from the previous stage as an input, then predicts features, drugs, or output depending on the network. Since learning is divided into multiple stages, which represent logical groupings, such as recovery periods in terms of time, using this approach allows obtaining feature and drug predictions solely based on the features and drugs of the previous step, allowing for more targeted learning. The *Feature* network essentially performs linear regression fitting over the stages, while *Drug* and *Output* networks perform logistic regression fitting, predicting probabilities as a result of sigmoid activation. Although we deal with 3 stages, this process may iteratively continue for i stages until the last stage, where the features and drugs at stage $n-1$ are combined and trained in the *Output* network to predict the output. The pipeline can be seen in Fig. 1. The parameters can be observed in Table 3.

Table 3. Pipeline parameters.

Parameters	**Values**
Folds	5
Optimizer	Adamax
Loss Function	MSE Loss
Epochs	50
Batch Size	100
Learning Rate	0.01

We train the full pipeline using 5-fold cross-validation with shuffled observations, however, shuffled observation IDs are still the same across stages to make sure that patient's result is trained against the results from other stages. Such validation means that 80% of the observations are held for training, while the remaining 20% is for validation, and this is performed 5 times. The parameters of the main network are reset for each fold. For all three networks, we use an *MSE* loss function, which provides the best learning results. In a single forward pass, which for each stage s predicts outputs for stage $s+1$, including the final stage, we use an Adamax optimizer that simultaneously optimizes the parameters of the networks for all three stages. We evaluate learning using *recall* and *F1-scores* on drug prediction in Stages 2 and 3 and on output in Stage 3. The results are presented as averages of folds for drugs and output, and each fold average is an average of batches.

Table 4. Multi-stage evaluation results on test folds for a given run.

	Drug Prediction Results		**Output Prediction Results**
Metrics	Stage 2	Stage 3	Output
Recall (fold average)	99.09%	99.98%	99.41%
F1-score (fold average)	98.87%	94.55%	66.37%

5. Results

Model training results can be seen in Fig. 2. We see drastically decreasing loss when predicting drugs and more oscillating loss when predicting the output. *T1-T2* means we predict timestep 2 drugs using timestep 1 input data, and *T2-T3* means we predict timestep 3 drugs using timestep 2 input data. The test evaluation results can be seen in Table 4. The model can predict drugs across stages with significantly high accuracy. There is room for improvement in output network prediction. The results validate the model training results shown in Fig. 2, where the learning of the output network is not very smooth. It should be noted that evaluation results may change depending on the run as a result of cross-validation; however, the results should be close across the runs.

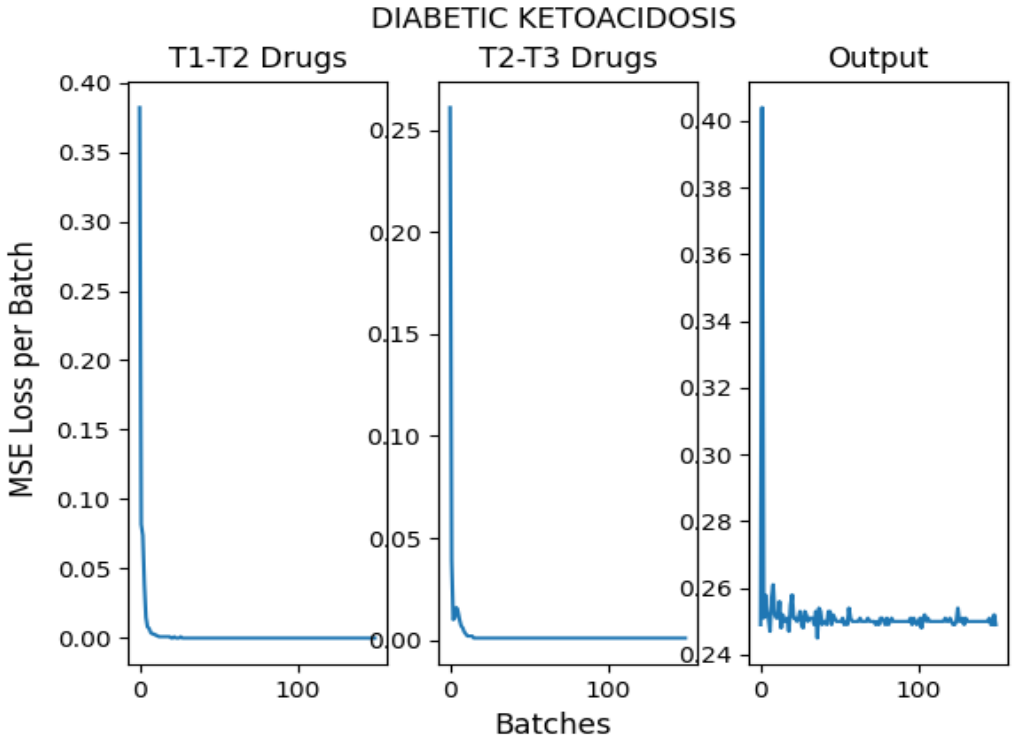

Fig. 2. Training results.

6. Limitations

We identify one main limitation of the paper. Stage-based and generated synthetic datasets do not allow us to benchmark our approach and results with existing similar studies. Although we provide evaluation results for these datasets, verifying the validity of the approach with existing studies could add significant value to the work. Providing benchmark datasets with the logic described in the paper can be part of future work.

7. Conclusion

In this work, we proposed an FNN-based pipeline that combines ensemble learning and iterative classification for modeling a multi-stage drug prescription procedure. The goal of

the approach was to make treatment assignments a dynamic rather than a one-stage static process by utilizing multiple predictor networks. Preprocessed and synthetic datasets for each stage are also provided. In addition, we also evaluated the performance of the whole approach based on how well the pipeline predicted drugs and the output on the testing folds. Although the results are promising, we acknowledge that the evaluations are based on synthetically derived datasets, which may affect the findings.

As part of future work, we can try to achieve class balance for bigger datasets, provide dataset benchmarks, improve the output prediction network, and come up with other logical groupings of a stage besides time. We also acknowledge that treatment predictions may become credible only after proper medical testing and validation.

References

[1] K. Gishyan, "Drug-treatment generation combinatorial algorithm based on machine learning and statistical methodologies", *Open Journal of Applied Sciences*, vol. 13, no. 4, pp.548–561, 2023.

[2] K. Gishyan, H. Sahakyan and L. Ananyan, "Time-stage driven pathfinding framework for optimized medical treatments", *Cogent Engineering*, vol. 10, no. 1, pp.2249258, 2023.

[3] Y. Ren and L. Zhang and P. H. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions", *IEEE Computational Intelligence Magazine*, vol. 11,no. 1,pp. 41–53, 2016.

[4] T. Dietterich, "Ensemble methods in machine learning", *International Workshop on Multiple Classifier Systems*, pp. 1–15, 2000.

[5] E. Ke Tang, P.N. Suganthan and X.Yao, "An analysis of diversity measures", *Machine Learning*, vol. 65,no. 1, pp. 247–271, 2006.

[6] Y. Ren, P.N. Suganthan, N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 236-244, 2014.

[7] A. Chandra and X. Yao, "Multi-objective ensemble construction, learning and evolution, *In Proc. PPSN Workshop Multi-objective Problem Solving from Nature (Part 9th Int. Conf. Parallel Problem Solving from Nature: PPSN-IX)*, pp. 913, 2006.

[8] M. Asafuddoula, B. Verma and Zh. Mengjie, "A divide-and-conquer-based ensemble classifier learning by means of many-objective optimization", *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 5, pp. 762-777, 2018.

[9] X. Zeng, W. Ouyang and X. Wang, "Multi-stage Contextual Deep Learning for Pedestrian Detection", *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 121–128, 2013.

[10] E. Alpaydin and C. Kaynak "Cascading classifiers", *Kybernetika*, vol. 34, no. 4, pp. 369–374, 1998.

[11] Zh. Sun, Y. Wang, T. Tan, and J. Cui, "Improving iris recognition accuracy via cascaded classifiers", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 435–441, 2005.

[12] Y. Wang and J. Patrick, "Cascading classifiers for named entity recognition in clinical notes", *Proceedings of the Workshop on Biomedical Information Extraction*, pp. 42–49, 2009.

[13] M. Sabokrou, M. Fayyaz, M. Fathy and R. Klett, "Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes", vol. 26, no. 4, pp. 1992–2004, 2017.

[14] Md. S. H. Shovon, Md. J. Islam, M. N. Ali Khan Nabil, Md Mohimen Molla, Akinul Islam Jony and MF Mridha, "Strategies for enhancing the multi-stage classification performances of her2 breast cancer from hematoxylin and eosin images", *Diagnostics*, vol. 12, no. 11, pp. 2825, 2022.

[15] I. I. Esener, S. Ergin, T. Yuksel, et al., "A new feature ensemble with a multistage classification scheme for breast cancer diagnosis", *Journal of Healthcare Engineering*, vol. 2017, doi:10.1155/2017/3895164, 2017.

[16] Y. Isler, A. Narin, M. Ozer, and M. Perc, "Multi-stage classification of congestive heart failure based on short-term heart rate variability", *Chaos, Solitons & Fractals*, vol. 118, pp. 145–151, 2019.

[17] Y. Kutlu and D. Kuntalp"A multi-stage automatic arrhythmia recognition and classification system", *Computers in Biology and Medicine*, vol. 41, no. 1, pp. 37–45, 2011.

K.R. Kruthika (Research Scholar), Rajeswari (Professor), H.D. Maheshappa (Professor),

[18] K.R. Kruthika, Rajeswari, H. D. Maheshappa, "Multistage classifier-based approach for Alzheimer's disease prediction and retrieval", *Informatics in Medicine Unlocked*, vol. 14, pp. 34–42, 2019.

[19] S. Manimurugan, S. Almutairi, M. M. Aborokbah, C. Narmatha, S. Ganesan, N. Chilamkurti, R. A. Alzaheb and H. Almoamari, "Two-stage classification model for the prediction of heart disease using IoMT and artificial intelligence', *Sensors*, vol. 22, no. 2, pp. 476, 2022.

[20] P. Tang, X. Yan, Y. Nan, Sh. Xiang, S. Krammer and T. Lasser, "FusionM4Net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification", *Medical Image Analysis*, vol. 76, pp. 102307, 2022.

[21] GB Praveen and A. Agrawal, "Multi stage classification and segmentation of brain tumor", *Proceedings International Conference on Computing for Sustainable Global Development (INDIACom)*,pp. 1628–1632, 2016.

[22] N. Sambyal, P. Saini, R. Syal and V. Gupta,"Aggregated residual transformation network for multistage classification in diabetic retinopathy", *International Journal of Imaging Systems and Technology*,vol. 31, no. 2, pp. 741–752, 2021.

[23] J. Oh, Ch. Park, H. Lee, B. Rim, Y. Kim, M. Hong, J. Lyu, S. Han and S. Choi, "OView-AI supporter for classifying pneumonia, pneumothorax, tuberculosis, lung cancer chest X-ray images using multi-stage superpixels classification", *Diagnostics*, vol. 13, no. 9, pp. 1519, 2023.

[24] N. Patki and R. Wedge and K. Veeramachaneni, "The synthetic data avult", *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 399-410,

[25] L.Xu, M. Skoularidou, A. Cuesta-Infante and K. Veeramachaneni, "Modeling tabular data using conditional gan", *Advances in Neural Information Processin Systems*, vol. 32, 2019.

Փուլային դասակարգման սխեմա՝ օպտիմալ բուժում ստանալու նպատակով

Կարեն Մ. Գիշյան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
e-mail: karengishyan.res@outlook.com

Ամփոփում

Գոյություն ունեցող մեքենայական ուսուցման և խորը ուսուցման պարամետրերում դասակարգման և ռեգրեսիայի կանխատեսման խնդիրները կարող են ներկայացվել որպես գործընթաց, որտեղ մոդելի ելքը հիմնված է մեկ փուլային մուտքագրման վրա: Իրական կյանքի սցենարների մեծ մասում հիվանդի համար ցանկալի բժշկական վիճակին հասնելը կարող է ներառել դեղերի նշանակման խնդիրների դինամիկ լուծում՝ հիմնված մուտքային տվյալների վրա տարբեր փուլերում, որտեղ յուրաքանչյուր փուլ տրամաբանական խմբավորում է, ինչպիսիք են՝ ժամանակի բաժանումը, ինտենսիվ թերապիայի բաժանմունքում մնալը և այլն: Տվյալ փուլի տվյալները ներկայացնում են վերականգնման առաջընթաց և կարող են էապես տարբերվել նախորդ և ապագա փուլերի տվյալներից: Թեև մեկ մոդելը կարող է լուծել առաջադրանքը, ուսուցման բազմափուլ ընթացակարգը կարող է ավելի հարմար լինել: Այս խնդիրը լուծելու համար մենք առաջարկում ենք FNN-ի վրա հիմնված մոդելների միավորման մոտեցում՝ կանխատեսելու այն դեղամիջոցները, որոնք հիվանդը պետք է ստանա վերականգնման գործընթացի յուրաքանչյուր փուլում: Վերջնական բժշկական դուրսգրման վայրը կանխատեսվում է դեղերի և վիճակի տվյալների հաջորդական կանխատեսումների արդյունքում: Այս աշխատանքում մենք համատեղում ենք մոդելային միավորումը և բազմափուլ իտերատիվ ուսուցումը՝ օպտիմալ դեղերի նշանակման առաջադրանքը լուծելու համար՝ որպես ներդրում առկա գրականության մեջ:

Բանալի բառեր՝ Փուլային դասակարգում, բուժման օպտիմալացում, մոդելների միավորում, մեքենայական ուսուցում:

Многоступенчатая схема классификации для оптимизации медицинского лечения

Карен М. Гишян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
e-mail: karengishyan.res@outlook.com

Аннотация

В большинстве реальных сценариев достижение желаемого медицинского состояния пациента может включать динамическое решение задач назначения лекарств на основе входных данных на разных этапах, где каждый этап представляет собой логическую группировку, такую как разделение временных шагов, пребывание в отделении интенсивной терапии и т. д. Данные на данном этапе представляют собой прогресс восстановления и могут принципиально

отличаться от наборов данных с предыдущих и будущих этапов. Хотя одну модель может решить задачу, многоэтапная процедура обучения может оказаться более подходящей. Для решения этой задачи мы предлагаем основанный на FNN ансамблевый подход для прогнозирования лекарств, которые пациент должен получать на каждом этапе процесса выздоровления. Окончательное место медицинской выписки прогнозируется в результате последовательного прогнозирования препаратов и особенностей. В этой работе мы объединяем ансамбль моделей и многоэтапное итеративное обучение для решения задачи создания оптимального рецепта на лекарства в качестве вклада в существующую литературу.

Ключевые слова: Многоэтапная классификация, оптимизация лечения, ансамбль моделей, машинное обучение.

Mathematical Problems of Computer Science 60, 52–62, 2023.
doi: 10.51408/1963-0109

UDC 004.934

Building a Speaker Diarization System: Lessons from VoxSRC 2023

Davit S. Karamyan[1,2] and Grigor A. Kirakosyan[2,3]

[1]Russian-Armenian University, Yerevan, Armenia
[2]Krisp.ai, Yerevan, Armenia
[3]Institute of Mathematics of NAS RA, Yerevan, Armenia
e-mail: {dkaramyan, gkirakosyan}@krisp.ai

Abstract

Speaker diarization is the process of partitioning an audio recording into segments corresponding to individual speakers. In this paper, we present a robust speaker diarization system and describe its architecture. We focus on discussing the key components necessary for building a strong diarization system, such as voice activity detection (VAD), speaker embedding, and clustering. Our system emerged as the winner in the Voxceleb Speaker Recognition Challenge (VoxSRC) 2023, a widely recognized competition for evaluating speaker diarization systems.
Keywords: Speaker recognition, Speaker diarization, VoxSRC 2023.
Article info: Received 27 September 2023; sent for review 28 September 2022; received in revised form 14 November 2023; accepted 16 November 2023.

1. Introduction and Related Work

Speaker diarization (SD) is the process of dividing audio into segments according to the speaker's identity. It is the process of determining "who spoke when" in a multi-speaker audio signal. A typical SD system usually consists of several steps: (1) segment the input audio into speech segments using a Voice Activity Detector (VAD), (2) generate speaker segments from the speech segments by either using a uniform sliding window segmentation or by detecting speaker turns, (3) extract speaker embeddings for each of the speaker segment, (4) group the resulting speaker embeddings into clusters using clustering algorithms. Commonly used clustering algorithms include Spectral Clustering (SC) [1] and Agglomerative Hierarchical Clustering (AHC) [2].

Despite recent advancements in speaker diarization [3], several factors make solving SD task difficult:

- *Uniform speaker segmentation*: Long segments very likely contain speaker turn boundaries, while short segments carry insufficient speaker information.

- *Unknown number of speakers*: In general, both the identity of the speakers and the number of speakers are unknown beforehand.
- *Speaker talk time*: A speaker needs to talk long enough to be accurately detected.
- *Overlap speech*: Talking over each other or interrupting.
- *Background noise, room acoustics*: Environmental sounds and room conditions can interfere with speaker recognition.
- *Consisting of multiple steps*: The SD system involves several steps, each of which introducing some level of error.

Speaker change detection systems have been proposed to mitigate the uniform segmentation issue [4, 5]. These systems involve a dedicated model trained to detect the exact moment when speakers change. To deal with a trade-off between long and short segment lengths, a group of works employs multi-scale segmentation [6, 7]. They use multiple scales (segment lengths) and fuse the similarity scores between embeddings obtained from the results of each scale.

To address the overlap speech problem, the recently introduced target-speaker voice activity detection (TS-VAD) model [8] has attracted much interest due to its great success in challenging tasks such as VoxSRC [9, 10, 11] and DIHARD-III [12]. Based on the speaker profiles obtained from a clustering-based diarization, the TS-VAD system can estimate each speaker's frame-level voice activities to refine the initial clustering-based results.

A line of research aims to improve the performance of conventional clustering-based methods by enhancing either through methods like embedding refinement [13, 14] or by refining similarity scores among speaker embeddings [1]. In [15], the Teacher-Student approach was employed to increase the robustness of the speaker embedding extractor against different acoustic conditions.

An alternative line of research ([16, 17, 18]) tackles the segmentation and clustering modules jointly. These models are referred to as "end-to-end". End-to-end algorithms have demonstrated their effectiveness over traditional modular systems in controlled situations with a limited number of speakers. However, their performance suffers in real-world recordings with a larger number of speakers.

In this paper, we describe our clustering-based SD system[1] for the Diarization Task of the 2023 VoxCeleb Speaker Recognition Challenge (VoxSRC23)[2]. The proposed system consists of several sub-modules, such as voice activity detection, speaker embedding extraction, clustering, and overlap speech detection (OSD). Along with the description, we will outline how to build a strong speaker diarization system and give a detailed analysis of each method.

2. About the VoxSRC 2023 Challenge

The goal of the VoxSRC challenge is to probe how well current methods can recognize speakers from speech obtained 'in the wild'. The Voxconverse dataset [19] was used for the speaker diarization task. The VoxConverse dataset contains 74 hours of human conversation

[1] `http://mm.kaist.ac.kr/datasets/voxceleb/voxsrc/data_workshop_2023/reports/krisp_report.pdf`

[2] `http://mm.kaist.ac.kr/datasets/voxceleb/voxsrc/interspeech2023.html`

extracted from YouTube videos. The dataset is divided into a development set (20.3h, 216 recordings) and a test set (53.5h, 232 recordings). The number of speakers in each recording has a wide range of variety from 1 speaker to 21 speakers. The audio comprises a variety of noises, such as background music, laughter, and so on. It also contains a significant portion of overlapping speech from 0% to 30.1% depending on the recording. The primary metric for this task is the Diarization Error Rate (DER), which is the sum of three terms: false alarm (FA, incorrectly marking non-speech as speech), missed detection (MS, incorrectly marking speech as non-speech) and speaker confusion error rate (CER, assigning the wrong speaker ID within a speech region). A separate evaluation dataset (VoxSRC-23 Test) was used to establish the rankings on the leaderboard.

3. System Configuration

3.1 Voice Activity Detection

Voice Activity Detection is the process of identifying speech segments within an audio signal, serving as an essential initial phase for speaker diarization. We employ four different VAD models, each designed to capture various facets of the task.

3.1.1 GRU-Based VAD

We use a stack of 4 Gated Recurrent Unit (GRU) layers, incorporating layer normalization between each layer. The final dense layer with sigmoid activation is responsible for calculating the likelihood of speech occurrence. With this setup, we generate a probability score for every 30ms of speech. Values nearing 1, signify the presence of speech, whereas values closer to 0 suggest its absence. We use the Voxconverse dev set for training and the Voxconverse test set for validation.

3.1.2 NC-Based VAD

We adopt the Noise Cancellation (NC) model [20] to detect voice activity. First, we apply the NC model to remove any noise and non-speech signals from the original audio. Subsequently, for each 50ms interval, we calculate the energy of that interval and establish a threshold. If the energy level exceeds the threshold, we label the segment as speech; otherwise, it is categorized as non-speech. Additionally, we apply simple post-processing steps to obtain homogeneous speech activity segments. The architecture of the NC model is the same as the GRU-VAD architecture, with the exception that it generates a mask. This mask is subsequently applied to the input spectrogram and transformed into a waveform using the Inverse Fourier Transform.

3.1.3 ASR-Based VAD

Another approach to detecting voice activity segments involves making use of an Automatic Speech Recognition (ASR) model to generate timestamps at the level of individual words. We derive word-level timestamps by employing the Conformer-Medium checkpoint available in the NeMo[3] package. Similar to NC-based VAD, here we also apply post-processing steps to obtain homogeneous speech segments.

[3] `https://github.com/NVIDIA/NeMo`

3.1.4 Pyannote VAD

We also provide evaluation results for an open-source VAD model available in *pyannote* package [21]. Specifically, we employ the *pyannote.audio 2.1*[4] segmentation pipeline for computing the voice activity regions.

Table 1. Detection Error Rate of the VAD model on Voxconverse test set.

#Model	FA	MISS	Detection Error
GRU-based	2.59%	1.40%	3.99%
NC-based	2.83%	2.09%	4.92%
ASR-based	3.04%	1.74%	4.79%
Pyannote	2.01%	1.19%	3.20%
Fusion	2.02%	0.82%	2.84%

Table 1 shows that *NC-based* and *ASR-based* VAD models have inferior performance compared to systems trained under direct supervision. However, when we fuse these models using a majority vote, we achieve a reduction in detection error rate by 0.36%.

3.2 Speaker Embedding Extraction

Speaker embeddings are fixed-size vector representations from a speech signal that exclusively capture unique characteristics of the speaker's identity. Speaker embeddings are commonly used to classify and discriminate between different speakers.

A few publicly available speaker embedding models listed in Table 2 were compared with the corresponding performance results and the corresponding training datasets. Performance results are reported in equal error rates (EER), which is a standard metric used to evaluate speaker verification.

Table 2. Equal Error Rate values for different embedding extraction models evaluated on the Voxceleb test benchmark.

Embedding	EER	Training Datasets
TitaNet-Large[22]	0.68% Vox1-Clean	Voxceleb1+Voxceleb2, Fisher, Switchboard, Librispeech
TitaNet-Small[22]	1.08% Vox1-Clean	Voxceleb1+Voxceleb2, Fisher, Switchboard, Librispeech
RawNet3[23]	0.89% Vox1-O	Voxceleb1+Voxceleb2
ECAPA-TDNN[24]	0.80% Vox1-Clean	Voxceleb1+Voxceleb2

To increase the accuracy of speaker recognition and speaker diarization for noisy audios, we finetune TitaNet-Small with the Teacher-Student method [15] by adding L_2-regularization term to the AAM loss [25], between embeddings for augmented and non-augmented versions of the same audio utterance. We follow the fine-tuning steps presented in [15]. For

[4] `https://huggingface.co/pyannote/segmentation`

fine-tuning, we use the VoxCeleb1 [26] and VoxCeleb2 [27] datasets. By employing this approach, we achieved EER comparable to the pre-trained TitaNet-Small[5] model under normal conditions. However, the technique demonstrated superior performance in noisy conditions.

3.3 Clustering

Once computed, the speaker embeddings are grouped into clusters. We use two different clustering algorithms for SD. One method relies on spectral clustering and the other is based on agglomerative hierarchical clustering.

3.3.1 Spectral Clustering

Our SC-based diarization is similar to [15]. We perform multi-scale segmentation [7] and extract embeddings with different window and shift sizes. The affinity matrices are constructed using the cosine similarity between segment embeddings and are then fused into a single matrix (see Fig. 1). We further apply the following sequence of refinement operations on the affinity matrix A (see Fig. 2):

- *Row-wise Thresholding*: For each row, keep the *top-p* largest elements and set the rest to 0
- *Symmetrization*: $Y = \frac{1}{2}(A + A^T)$
- *Diffusion*: $Y = AA^T$

Afterwards, we apply the spectral clustering algorithm to obtain speaker IDs. The number of speakers is determined using the maximal eigen-gap approach [1].

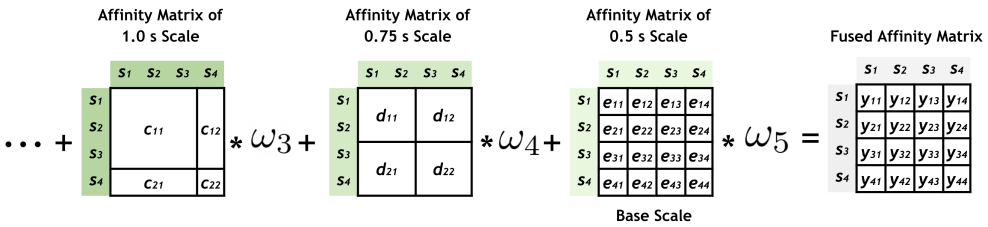

Fig. 1. Multi-scale segmentation scheme.

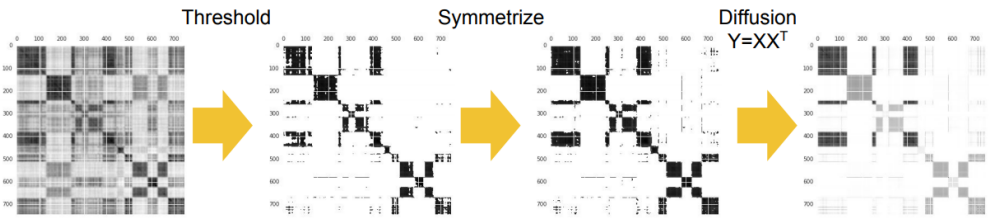

Fig. 2. Refinement operations on the affinity matrix.

[5] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/titanet_small

3.3.2 Agglomerative Hierarchical Clustering

First, we extract speaker embeddings from uniformly segmented speech regions. Then, we refine these embeddings through spectral dimensionality reduction[6] and affinity aggregation (AA) techniques [14]. We merged consecutive segments into a longer one if the distance was greater than the *segment threshold*. Afterwards, we perform a plain agglomerative clustering on the refined embeddings with a relatively high *stop threshold* to obtain the clusters with high confidence. The clusters from AHC were further processed using the short-duration filter [2, 10]. We categorize a cluster as "short" if the combined duration of that cluster is below the specified *duration threshold*. Later, each short cluster is assigned to the nearest long cluster based on the cosine distance of their central embeddings. Finally, if a short cluster significantly differs from all long clusters, which means that the distance between them is lower than a *speaker threshold*, we consider it as a new speaker.

3.4 Overlap Speech Detection

To detect regions where two or more speakers are speaking simultaneously, we use *pyannote overlap speech detection* pipeline[7]. After an overlapped region is detected, we replace the label with the two closest speakers near this region in the time domain.

3.5 Fusion

To improve the diarization accuracy, a series of studies were conducted on the fusion method of multiple diarization results. More recently, the diarization output voting error reduction (DOVER) method [28] was proposed to combine multiple diarization outputs based on the voting scheme. The DOVER method has an implicit assumption that there is no overlapping speech, i.e., at most only one speaker is assigned for each time index. To accommodate diarization outputs with overlapping speakers, the DOVER-LAP [29] method was subsequently introduced.

We combine different diarization systems using the DOVER-Lap[8] fusion method with the Hungarian label mapping algorithm.

4. Experimental Results

Table 3 shows the results on the voxconverse test set and the challenge evaluation test set. The first row of the table displays the baseline result (*VGG baseline*), provided by the challenge organizers. We start with the pyannote VoxSRC22 pipeline (#1) as our initial system and enhance it by applying the affinity aggregation technique (#2) to refine the embeddings. This adjustment results in a reduction of 0.59% in DER on the voxconverse test set.

Next, we designed several diarization systems based on spectral clustering with different embedding extractors ($\#3 - \#9$). These systems all rely on uniform speaker segmentation, which leads to speaker errors, mainly around the speaker turns. To mitigate this issue, we use

[6] `https://scikit-learn.org/stable/modules/generated/sklearn.manifold.SpectralEmbedding.html`

[7] `https://huggingface.co/pyannote/overlapped-speech-detection`

[8] `https://github.com/desh2608/dover-lap`

different segmentation setups by changing both the window size and the shift size. Multi-scale segmentation (#5, #8) is also designed to tackle this problem and to remove noisy entries from the affinity matrix. Furthermore, to make the systems more robust, we apply a sequence of refinement operations on the affinity matrix. In single-scale segmented setups, we establish the *top-p* value for row-wise thresholding as 8. In the case of multi-scale segmented setups, this value is adjusted to 30. As one can see from Table 3, multi-scale segmented systems outperform single-scale ones by a margin of 0.3%. Surprisingly, system #9, which was finetuned with the Teacher-Student technique, achieves a similar score (5.23%) on the voxconverse test set without using multi-scale segmentation.

As noted in [10], SC-based and AHC-based clustering methods complement each other. Through our experiments, we also observed similar behaviour. Spectral clustering provides a more precise estimation of the number of speakers, whereas AHC-based clustering tends to consistently overestimate it. Conversely, AHC-based clustering excels at identifying the dominant speakers and demonstrates superior performance on shorter audio files compared to spectral clustering. We conduct a hyperparameter search for AHC-based systems (#10, #11, #12) on the voxconverse test subset to determine the optimal values for *segment threshold*, *stop threshold*, *duration threshold*, and *speaker threshold*. As it is illustrated in Table 3, AHC-based systems show slightly worse DER scores (5.32%-5.41%) compared to SC-based systems.

Table 3. The performance of different speaker diarization systems.

N	**System**	**Window** [s]	**Shift** [s]	**Voxconverse Test**	**VoxSRC-23 Test**
				DER[%]	**DER**[%]
	VGG baseline	-	-	-	8.68
#1	Pyannote VoxSRC22	-	-	5.89	7.33
#2	Pyannote VoxSRC22+AA	-	-	5.30	-
#3	TitaNet-Large-SC	1.0	0.75	6.00	-
#4	TitaNet-Large-SC	2.0	1.0	5.59	-
#5	TitaNet-Large-SC	[2.0, 1.5, 0.75]	[1, 0.5, 0.25]	5.25	-
#6	ECAPA-TDNN-SC	1.0	0.75	6.05	-
#7	ECAPA-TDNN-SC	2.0	1.0	5.71	-
#8	ECAPA-TDNN-SC	[2, 1.5, 0.75]	[1, 0.5, 0.25]	5.38	-
#9	TitaNet-Small-SC	1.5	0.5	5.23	-
#10	TitaNet-Large-AHC	1.5	0.5	5.41	-
#11	ECAPA-TDNN-AHC	1.5	0.5	5.38	-
#12	RawNet3-AHC	1.5	0.75	5.32	-
	Fusion(3+4+5+6+7+8)+OSD	-	-	4.80	6.35
	Fusion(2+3+4+5+6+7+8)+OSD	-	-	4.76	5.98
	Fusion(2+5+8+9+10+11+12)+OSD	-	-	4.39	4.71

Our best system combines 7 different systems fused by DOVER-Lap. Among them, 3 systems are based on spectral clustering, while 4 systems are based on AHC (including pyannote system #2). We first fused the systems and then dealt with the overlap because fusing with overlapping labels did not demonstrate any improvement on the voxconverse test set. This fused system achieves 4.39% DER on the voxconverse test set and 4.71% DER on the challenge evaluation set, which ranks 2nd place in the VoxSRC 2023 challenge.

5. Discussions

Throughout our experiments, we observed that better performance on widely adopted speaker verification evaluation protocols does not lead to better diarization performance. Additionally, the embedding extractors did not encounter situations where multiple speakers were present in audio utterances. Such scenarios are unavoidable in speaker diarization due to factors like overlapping speech and speaker transitions.

In contrast to speaker verification, which uses speaker embeddings to represent an endless number of speakers, speaker diarization only uses embeddings to represent a small number of speakers in a single session. For instance, only a small part of the information included in the embeddings will be used to distinguish between a small number of speakers, even if high-dimensional embeddings are extracted.

Another drawback of conventional clustering-based SD systems is that they do not take into consideration embedding ordering. Conversations involving multiple speakers are highly structured, and turn-taking behaviours are not dispersed randomly throughout time.

In our future work, we plan to investigate speaker verification evaluation protocols that better simulate the diarization scenario. Additionally, we will explore techniques aimed at adapting and contextualizing speaker embeddings for the speaker diarization task, as well as exploring approaches to leverage ordering information of embeddings.

6. Conclusions

In this paper, we described our submitted SD system for the diarization task of the 2023 VoxSRC challenge. We mainly focused on reducing speaker confusion errors. To achieve this goal, we used various methods, such as multi-scale segmentation, affinity refinement operations, and teacher-student techniques to make our SD systems robust to background noise and errors that might arise from uniform speech segmentation. Our final system yielded notable results, reaching a DER of 4.39% on the voxconverse test set and 4.71% on the challenge evaluation set.

References

[1] Q. Wang, C. Downey, L. Wan, P. Mansfield and I.Moreno, "Speaker diarization with LSTM", *2018 IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 5239-5243, 2018.

[2] X. Xiao, N. Kanda, Z. Chen, T. Zhou, T. Yoshioka, S. Chen, Y. Zhao, G. Liu, Y. Wu, J. Wu and et.a, "Microsoft speaker diarization system for the voxceleb speaker recognition challenge 2020", *IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP)*, pp. 5824-5828, 2021.

[3] T. Park, N. Kanda, D. Dimitriadis, K. Han, S. Watanabe and S. Narayanan, "A review of speaker diarization: Recent advances with deep learning", *Computer Speech & Language*, vol. 72, pp. 101317, 2022.

[4] R. Yin, H. Bredin and C. Barras, "Speaker change detection in broadcast TV using bidirectional long short-term memory networks", *Interspeech 2017*, 2017.

[5] W. Xia, H. Lu, Q. Wang, A. Tripathi, Y. Huang, I. Moreno and H. Sak, "Turn-to-diarize: Online speaker diarization constrained by transformer transducer speaker turn detection", *IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 8077-8081, 2022.

[6] T. Park, M. Kumar and S. Narayanan, "Multi-scale speaker diarization with neural affinity score fusion", *IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP)*, pp. 7173-7177, 2021.

[7] Y. Kwon, H. Heo, J. Jung, Y. Kim, B. Lee and J. Chung, "Multi-scale speaker embedding-based graph attention networks for speaker diarisation", *IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 8367-8371, 2022.

[8] I. Medennikov, M. Korenevsky, T. Prisyach, Y. Khokhlov, M. Korenevskaya, I. Sorokin, T. Timofeeva, A. Mitrofanov, A. Andrusenko, I. Podluzhny , A. Laptev and A. Pomanenko, "Target-Speaker Voice Activity Detection: a Novel Approach for Multi-Speaker Diarization in a Dinner Party Scenario", arXiv:2005.07272, 2020.

[9] W. Wang, D. Cai, Q. Lin, L. Yang, J. Wang, J. Wang and M. Li, "The dku-dukeece-lenovo system for the diarization task of the 2021 voxceleb speaker recognition challenge", arXiv:2109.02002, 2021.

[10] W. Wang, X. Qin, M. Cheng, Y. Zhang, K. Wang, and M. Li, "The dku-dukeece diarization system for the voxceleb speaker recognition challenge 2022', arXiv:2210.01677, 2022.

[11] M. Cheng, W. Wang, Y. Zhang, X. Qin and M. Li, "Target-speaker voice activity detection via sequence-to-sequence prediction" *IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 1-5, 2023.

[12] Y. Wang, M. He, S. Niu, L. Sun, T. Gao, X. Fang, J. Pan, J. Du and C. Lee, "USTC-NELSLIP system description for DIHARD-III challenge", arXiv:2103.10661, 2021.

[13] J. Wang, X. Xiao, J. Wu, R. Ramamurthy, F. Rudzicz and M. Brudno, "Speaker diarization with session-level speaker embedding refinement using graph neural networks", *IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 7109-7113, 2020.

[14] Y. Kwon, J. Jung, H. Heo, Y. Kim, B. Lee and J. Chung, "Adapting speaker embeddings for speaker diarisation", arXiv:2104.02879, 2021.

[15] D. Karamyan, G. Kirakosyan and S. Harutyunyan, "Making speaker diarization system noise tolerant", *Mathematical Problems of Computer Science*. vol.59, pp. 57-68, 2023.

[16] A. Zhang, Q. Wang, Z. Zhu, J. Paisley and C. Wang, "Fully supervised speaker diarization", *Proceedings of IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*, pp. 6301-6305, 2019.

[17] Y. Fujita, N. Kanda, S. Horiguchi, K. Nagamatsu and S. Watanabe, "End-to-end neural speaker diarization with permutation-free objectives", *Proc. Interspeech*, pp. 4300-4304, 2019.

[18] Y. Fujita, N. Kanda, S. Horiguchi, Y. Xue, K. Nagamatsu and S. Watanabe, "End-to-end neural speaker diarization with self-attention", *Proceedings of IEEE Automatic Speech Recognition And Understanding Workshop (ASRU)*, pp. 296-303, 2019.

[19] J. Chung, J. Huh, A. Nagrani, T. Afouras and A. Zisserman, "Spot the conversation: Speaker diarisation in the wild", *Proc. Interspeech 2020*, pp. 299-303, 2020.

[20] Y. Xu, J. Du, L. Dai, and C. Lee, "A regression approach to speech enhancement based on deep neural networks", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 7-19, 2014.

[21] Bredin, H., Yin, R., Coria, J., Gelly, G., Korshunov, P., Lavechin, M., Fustes, D., Titeux, H., Bouaziz, W. & Gill, M. Pyannote. audio: neural building blocks for speaker diarization. *ICASSP 2020-2020 IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*. pp. 7124-7128 (2020).

[22] N. Koluguri, T. Park and B. Ginsburg, "TitaNet: Neural model for speaker representation with 1D Depth-wise separable convolutions and global context", *Proceedings of IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP)*, pp. 8102-8106, 2022.

[23] J. Jung, Y. Kim, H. Heo, B. Lee, Y. Kwon and J. Chung, "Pushing the limits of raw waveform speaker recognition", *23rd Annual Conference of the International Speech Communication Association, INTERSPEECH 2022*, pp. 2228-2232, 2022.

[24] B. Desplanques, J. Thienpondt and K. Demuynck, "ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification", *Proc. Interspeech 2020*, pp. 3830-3834, 2020.

[25] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690-4699, 2019.

[26] A. Nagrani, J. Chung and A. Zisserman, "Voxceleb: a large-scale speaker identification dataset", arXiv:1706.08612, 2017.

[27] J. Chung, A. Nagrani and A. Zisserman, "Voxceleb2: Deep speaker recognition", arXiv :1806.05622, 2018.

[28] A. Stolcke and T. Yoshioka, "DOVER: A method for combining diarization outputs", *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 757-763, 2019.

[29] D. Raj, L. Garcia-Perera, Z. Huang, S. Watanabe, D. Povey, A. Stolcke and S. Khudanpur, "Dover-lap: A method for combining overlap-aware diarization outputs", *IEEE Spoken Language Technology Workshop (SLT)*, pp. 881-888, 2021.

Խոսնակների դիարիզացիայի համակարգի կառուցում. դասեր VoxSRC-ից

Դավիթ Ս. Քարամյան[1,2] և Գրիգոր Ա. Կիրակոսյան[2,3]

[1]Հայ-Ռուսական համալսարան, Երևան, Հայաստան
[2]Krisp.ai, Երևան, Հայաստան
[3]ՀՀ ԳԱԱ մաթեմատիկայի ինստիտուտ,Երևան, Հայաստան
e-mail: {dkaramyan, gkirakosyan }@krisp.ai

Ամփոփում

Խոսնակների դիարիզացիայի նպատակը աուդիո ձայնագրության մեջ տարբեր խոսնակների հայտնաբերելն ու առանձնացնելն է: Այս հոդվածում ներկայացված է դիարիզացման հուսալի համակարգ, ինչպես նաև նկարագրված են այդ համակարգի կառուցվածքն ու հիմնական բաղադրիչները, ինչպիսիք են ձայնի հայտնաբերումը, խոսնակների ձայնային հատկանիշներ դուրս բերող մոդելը և կլաստերացումը, որոնք անհրաժեշտ են դիարիզացման հուսալի համակարգ ստեղծելու համար: Այս համակարգը հաղթող է ճանաչվել Voxceleb Speaker Recognition Challenge (VoxSRC) 2023 մրցույթում, որը լայնորեն ճանաչված է խոսնակների դիարիզացման համակարգերի գնահատման մրցույթում:

Բանալի բառեր՝ Խոսնակների նույնականացում, խոսնակների դիարիզացիա, VoxSRC2023:

Построение системы диаризации дикторов: опыт из VoxSRC 2023

Давид С. Карамян[1,2] и Григор А. Киракосян[2,3]

[1]Российско-Армянский университет, Ереван, Армения
[2]Krisp.ai, Ереван, Армения
[3]Институт математики НАН РА, Ереван, Армения
e-mail: {dkaramyan, sharutyunyan, gkirakosyan}@krisp.ai

Аннотация

Диаризация дикторов - это процесс разделения аудиозаписи на сегменты, которые соответствуют отдельным дикторам. В этой статье представлена надежная система диаризации говорящих и описана архитектура данной системы. Сосредоточено внимание на обсуждении ключевых компонентов, таких как обнаружение речевой активности экстрактор речевых характеристик и кластеризация, которые необходимы для создания надежной системы диаризации. Данная система стала победителем конкурса Voxceleb Speaker Recognition Challenge (VoxSRC) 2023, широко признанного конкурса по оценке систем диаризации дикторов.

Ключевые слова: Распознавание по голосу, диаризация дикторов, VoxSRC 2023.

Mathematical Problems of Computer Science 60, 63–71, 2023.
doi: 10.51408/1963-0110

UDC 004.93

Convolutional Neural Network (CNN) Layer Development for Effectiveness of Classification Tasks

Rafayel M. Veziryan[1] and Rafayel N. Khachatryan[2]

[1]Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
[2]Questrade Armenia Inc., Yerevan, Armenia
e-mail: rafaelveziryan@gmail.com, raf.khachatryan4@gmail.com

Abstract

This paper presents a novel 2D convolutional layer motivated by the principles of Partial Differential Equation (PDE) of Neural Interaction. Our objective is to leverage this layer to enhance the classification accuracy of Deep Convolutional Neural Networks (DCNN) for various classification tasks. We place a particular emphasis on its integration within the ResNet architecture, and we conduct experimental evaluations on the CIFAR10 and STL10 datasets to validate its efficacy.
Keywords: DCNN, Classification task, Image processing, ResNet, PDE, Cable equation, Grid method.
Article info: Received 30 September 2023; sent for review 2 October 2023; received in revised form 14 November 2023; accepted 16 November 2023.
Acknowledgement: We thank Professor Armenak Babayan for proposing the fundamental problem that guided this research, and for his consistent consultations.

1. Introduction

Deep Learning is a subfield of machine learning [1]. Neural networks simulate the learning process of the human brain. This article explores the synergy between deep learning and PDEs, specifically in the context of enhancing image classification accuracy using ResNet architecture [2].

PDEs hold significant importance in the realms of sound, image, and video processing [3]. Their application in image processing primarily revolves around noise removal and reconstruction[4, 5]. The foundational models relying on PDEs are adept at noise reduction while simultaneously maintaining the image integral features.

Transitioning our focus to DCNNs, challenges like gradient vanishing and gradient exploding have often been obstacles in training Neural Networks effectively. ResNet, or Residual Network, addresses such issues. At its core, the ResNet architecture employs Residual Blocks. These blocks incorporate skip connections, bypassing selected layers, offering a unique approach to handling these challenges. We will delve deeper into the intricacies of the ResNet architecture in the ensuing sections.

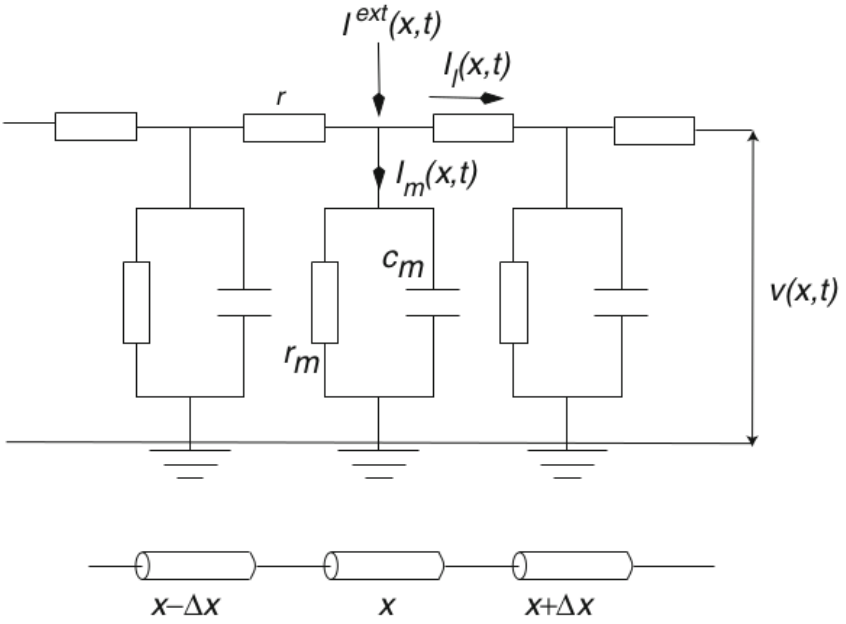

Fig. 1. Equivalent scheme for successive cylindrical segments of a dendritic membrane of a neural cell.

The essence of this article lies in optimizing DCNNs, specifically ResNet, by leveraging the capabilities of PDEs. An insightful source for our research is Bresslof's book, which introduces the Cable Equation [6](Section 1.4). The cable equation describes the variation of potential in neural cells.

Our principal objective is to harness the prowess of PDEs in image processing and integrate it within DCNNs, particularly with ResNet.

2. Mathematical Background

The main focus of this article is on the application of a CNN layer, obtained from Cable equation for classification tasks. As it was mentioned, the cable equation is an important mathematical model of the potential transmission in neural cells. The equivalent scheme is described in Fig. 1.

Taking into account that the transmission of potentials, that is, the transmission of information in neuronal cells, is governed by the equation described above, we propose to use the discretization of this equation in the construction of a neural network. According to the Book designations, the potential transmission in a neuron cell is described by the following equation:

$$\tau_m \frac{\partial v(x,t)}{\partial t} = -v(x,t) + \lambda_m^2 \frac{\partial^2 v(x,t)}{\partial x^2} + r_m I_{ext}(x,t), \;\; t \geq 0, \tag{1}$$

where v is a membrane potential at position x along a cable at time t, C_m is a capacity per unit of the membrane, R is a resistance of the intracellular fluid, R_m is a cell membrane resistance, a is a cable radius, $\tau_m = R_m C_m$ is a membrane time constant and $\lambda_m = (R_m a/2R)^{1/2}$ is a membrane space constant.

Therefore, we propose to apply the discretization of this equation in a neural network. We will employ the grid method to discretize this equation. The grid method is a fundamental technique that transforms PDEs into discrete computational forms. PDEs are converted into algebraic equations by segmenting space into a grid of discrete points, facilitating integration. The distribution of these points establishes the foundation upon which the PDEs are approximated using the finite difference method. The discretized form of the formula is obtained through the finite difference method.

In schematic form, Equation (1) takes the following form:

$$\frac{\partial u}{\partial t} = \alpha \Delta u - u$$

or

$$u_t = \alpha(u_{xx} + u_{yy}) - u.$$

Replace the partial derivatives with their finite difference approximations

$$\frac{u_{i,k}^{t+1} - u_{i,k}^{t}}{\tau} = \alpha \frac{u_{i-1,k}^{t} - 2u_{i,k}^{t} + u_{i+1,k}^{t}}{h_x^2} + \alpha \frac{u_{i,k-1}^{t} - 2u_{i,k}^{t} + u_{i,k+1}^{t}}{h_y^2} - u_{i,k}^{t}. \tag{2}$$

From (2) follows

$$u_{i,k}^{t+1} = u_{i,k}^{t}(1-\tau) + \frac{u_{i-1,k}^{t} - 2u_{i,k}^{t} + u_{i+1,k}^{t}}{\Phi^2} + \frac{u_{i,k-1}^{t} - 2u_{i,k}^{t} + u_{i,k+1}^{t}}{\Psi^2}, \tag{3}$$

where

$$\Phi^2 = \frac{h_x^2}{\alpha \cdot \tau}, \quad \Psi^2 = \frac{h_y^2}{\alpha \cdot \tau}.$$

The resulting scheme is called an explicit scheme because the solution's value at the given moment $t+1$ is strictly obtained by the solution's value in the previous t moment.
Equation (3) is equivalent to

$$u_{i,k}^{t+1} = (1-\tau) \cdot u_{i,k}^{t} + (P_1 + P_2) \cdot U, \tag{4}$$

where

$$U = \begin{bmatrix} u_{i-1,k-1}^{t} & u_{i-1,k}^{t} & u_{i-1,k+1}^{t} \\ u_{i,k-1}^{t} & u_{i,k}^{t} & u_{i,k+1}^{t} \\ u_{i+1,k-1}^{t} & u_{i+1,k}^{t} & u_{i+1,k+1}^{t} \end{bmatrix} sP_1 = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{\Phi^2} & -\frac{2}{\Phi^2} & \frac{1}{\Phi^2} \\ 0 & 0 & 0 \end{bmatrix} P_2 = \begin{bmatrix} 0 & \frac{1}{\Psi^2} & 0 \\ 0 & -\frac{2}{\Psi^2} & 0 \\ 0 & \frac{1}{\Psi^2} & 0 \end{bmatrix}$$

P_1 and P_2 are defined as two-dimensional, weighted convolution operators for the neural network with weights Φ, Ψ, and τ is also a weight. (3) represents our CNN layer, $u_{i,k}^{t+1}$ is our present layer, and $u_{i,k}^{t}$ is our previous layer, which we will call a cable equation layer.

3. Architecture

In this paper, we introduce a new component within DCNN, which we call a cable equation layer, and it is designed to be trainable. Our new layer usage highlights the features of the image, and since the layer is trainable, it allows us to optimize its configuration for

image processing. Traditional DCNNs begin with a standard convolutional layer. Here we introduce our cable equation layer at the outset, which enables us to process the image before the other layers of the network process it further. We investigate the efficacy of our layer in two convolutional blocks: first, we integrate the cable equation layer into a standard CNN Block, and second, we integrate the cable equation layer into the ResNet Block (see Fig. 2).

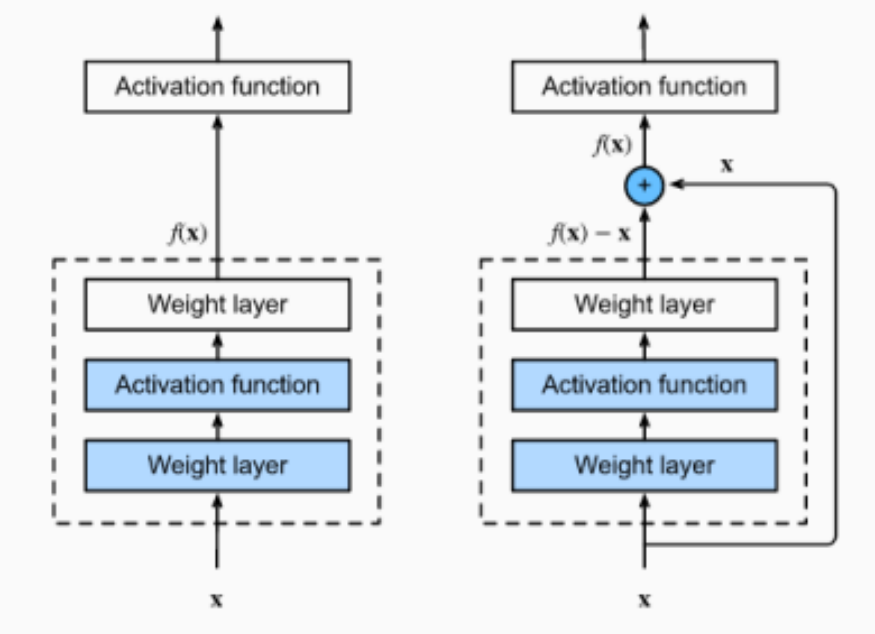

Fig. 2. Left: Standard CNN block, Right: ResNet block.

4. Experiments

The CIFAR10 dataset [7] is used for our experiments, consisting of 50K training images and 10K testing images. Input images have 3 channels and a size of 32×32, and use the STL10 dataset[8], which consists of 5K training images and 8K testing images, and have 3 channels and a size of 96×96. Network architecture will be presented in the table. Our Cable Equation layer incorporates BatchNorm [9] and employs the ReLU activation function. We conducted two experiments incorporating the Cable Equation layer. In the first experiment, we introduced the Cable Equation layer along with the preceding layers before the core Neural Network. For the second experiment, we crafted a residual block infused with a Cable Equation layer, iteratively applied one or more times, strategically positioned prior to the Neural Network. Table 1 below is a description of the architecture of the original and experimental models for the CIFAR10 dataset.

Initially, we implement the k-cable equalization layer, followed by 3x3 convolutions generating 64 output channels. Subsequently, the architecture encompasses 8 ResNet Blocks, partitioned into four sections, each with output channels (64, 128, 256, 512). The sequence continues with 4x4 averaging for CIFAR10 and 12x12 for STL10, a fully connected layer, and culminates with LogSoftmax activation. We used this architecture for both datasets.
For the CIFAR10 dataset, we initiate preprocessing steps before feeding the data into the neural network. This involves padding each side with 4 pixels, followed by a random crop of size 32x32. Additionally, a RandomHorizontalFlip operation is applied, and ultimately,

Table 1: CIFAR10 dataset: From top to bottom: Original ResNet, ResNet with 1, 2, 3 CabEq standard blocks, ResNet with 1, 2, 3 CabEqBl Residual Blocks. O. S. is the output shape after each layer.

Original							
CabEqSdBl1	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 1$						
CabEqSdBl2	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 2$						
CabEqSdBl3	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 3$						
CabEqRNBl1	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 1$	$\begin{bmatrix} 3\times 3, \\ 64 \end{bmatrix}$	$\begin{bmatrix} 3\times 3, \\ 64 \\ 3\times 3, \\ 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 128 \\ 3\times 3, \\ 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 256 \\ 3\times 3, \\ 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 512 \\ 3\times 3, \\ 512 \end{bmatrix} \times 2$	Average pooling(4×4), 10d FC, LogSoftmax
CabEqRNBl2	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 2$						
CabEqRNBl3	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 3$						
O. S.	32×32	32×32	32×32	16×16	8×8	4×4	1×1

the image is normalized using the means of (0.4914, 0.4822, 0.4465) and standard deviations of (0.2023, 0.1994, 0.2010). These transformations are conducted on the training data. For the test data, normalization is exclusively applied using the same means and standard deviations. The outcomes for CIFAR10 are presented in Table 2.

Table 2: The result of testing CIFAR10 dataset in Original ResNet and ResNet with Cable equation standard blocks and Cable equation ResNet blocks.

Model	Accuracy(%)	Parameter cnt
Original	88.92	11181642
CabEqSdBl 1	89.41	11181675
CabEqSdBl 2	89.57	11181708
CabEqSdBl 3	89.4	11181741
CabEqRNBl 1	89.68	11181708
CabEqRNBl 2	89.83	11181774
CabEqRNBl 3	88.9	11181840

For the STL10 dataset, we employ preprocessing procedures tailored to its distinct characteristics. To prepare the data for neural network input, we apply 4-pixel padding to all sides, followed by a randomized crop of dimensions 96x96. Additionally, a RandomHorizontalFlip operation is implemented. Subsequently, the image is normalized using mean values of (0.44671097, 0.4398105, 0.4066468) and standard deviations of (0.2603405, 0.25657743, 0.27126738). These transformations are integral to the training data, while for the test data, normalization is consistently applied using the same mean and standard deviation parameters. Table 3 shows the architecture for STL10. The outcomes for STL10 are presented in Table 4.

Table 3: STL10 dataset: From top to bottom: Original ResNet, ResNet with 1, 2, 3 CabEq layers, ResNet with 1, 2, 3 CabEq Residual Blocks. O. S. is the output shape after each layer.

Original							
CabEqSdBl1	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 1$						
CabEqSdBl2	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 2$						
CabEqSdBl3	$\begin{bmatrix} 3\times 3, & 3 \end{bmatrix} \times 3$						
CabEqRNBl1	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 1$	$\begin{bmatrix} 3\times 3, \\ 64 \end{bmatrix}$	$\begin{bmatrix} 3\times 3, \\ 64 \\ 3\times 3, \\ 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 128 \\ 3\times 3, \\ 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 256 \\ 3\times 3, \\ 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, \\ 512 \\ 3\times 3, \\ 512 \end{bmatrix} \times 2$	Average pooling(12×12), 10d FC, LogSoftmax
CabEqRNBl2	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 2$						
CabEqRNBl3	$\begin{bmatrix} 3\times 3, & 3 \\ 3\times 3, & 3 \end{bmatrix} \times 3$						
O. S.	96×96	96×96	96×96	48×48	24×24	12×12	1×1

Table 4: The result of testing STL10 dataset in Original ResNet and ResNet with Cable equation standard blocks and Cable equation ResNet blocks.

Model	Accuracy(%)	Parameter cnt
Original	73.625	11181642
CabEqSdBl 1	75.68	11181675
CabEqSdBl 2	75.33	11181708
CabEqSdBl 3	70.8	11181741
CabEqRNBl 1	75.025	11181708
CabEqRNBl 2	75.6625	11181774
CabEqRNBl 3	75.45	11181840

Our training strategy involves stochastic gradient descent (SGD)[10] coupled with a momentum of 0.9. The learning rate is managed using a OneCycleLR scheduler[11], featuring a maximum rate of 0.5. Throughout the training process, we iterate through 100 epochs for both datasets. Fig. 3 and Fig. 4 show the loss and accuracy for testing images per epoch for the CIFAR10 and STL10 datasets.

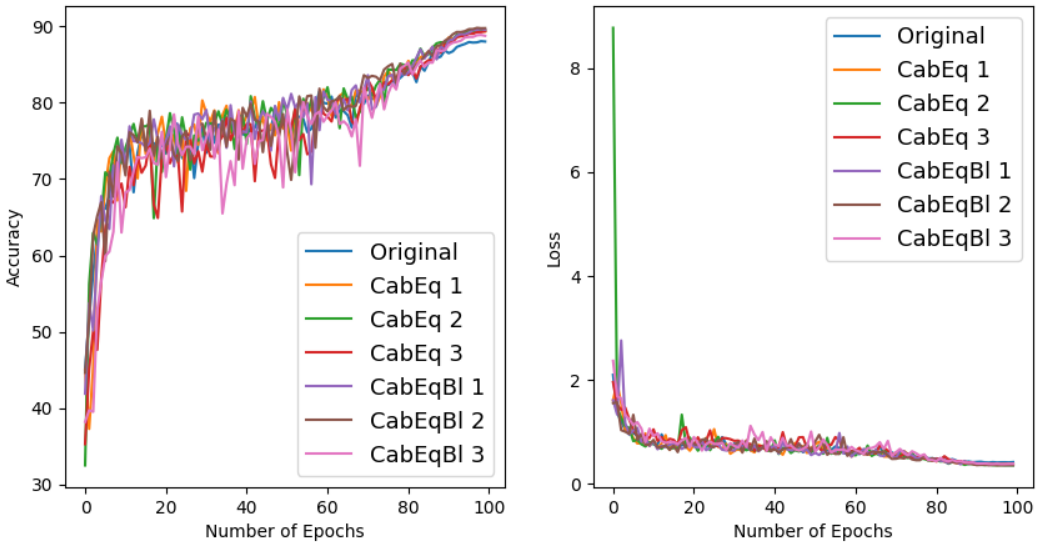

Fig. 3. Left: Relation between accuracy and epoch count for testing images of CIFAR10 dataset, Right: Relation between loss and epoch count for testing images of CIFAR10 dataset

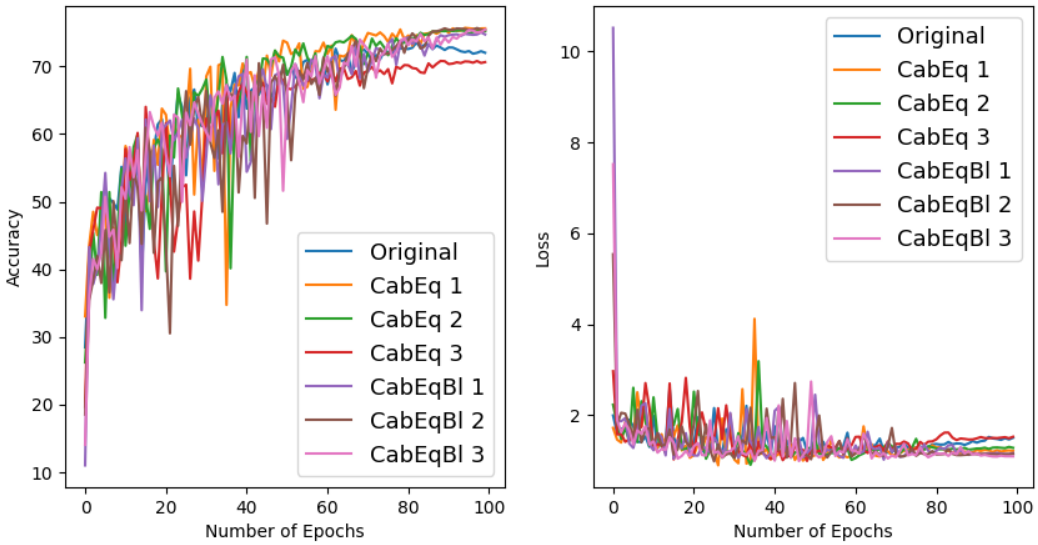

Fig. 4. Left: Relation between accuracy and epoch count for testing images of STL10 dataset, Right: Relation between loss and epoch count for testing images of STL10 dataset

5. Conclusion

The aim of this article is to increase the efficiency of DCNNs for classification tasks. We obtained the discretization of the PDE using the Grid method. Based on this, a convolution layer was constructed from the obtained convolution operators, which we called this the Cable equation layer. Motivated by the application of PDEs in image processing, we built the architecture by adding the Cable equation layer in front of the DCNN as a learnable

image processing layer. This enables us to find the best layer to process the image. Our experiments were conducted on the ResNet architecture, and the tests were performed on the CIFAR10 and STL10 datasets. Based on the obtained results, we can say that having a layer with a few parameters can increase effectiveness. The subject of further research can be integrating the Cable equation layer into other architectures, as well as thinking about creating new architectures based on the Cable equation layer itself.

References

[1] I. H. Sarker, "Deep Learning: A comprehensive overview on techniques, taxonomy, applications and tesearch directions", *SN Computer Science*, vol. 2, no. 6, 420, 2021.

[2] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016.

[3] K. Mikula, "Image processing with partial differential equations", *Modern Methods in Scientific Computing and Applications. NATO Science Series*, vol. 75, pp. 283-321, 2002.

[4] F. Guichard, L. Moisan and J.-M. Morel, "A review of P.D.E.models in image processing and image analysis", *Journal dePhysique IV (Proceedings)*, vol. 12, no. 1, pp. 137-154, 2002.

[5] L. Ruthotto and E. Haber, "Deep neural networks motivated by partial differential equations", *Journal of Mathematical Imaging and Vision*, vol. 62, no. 3, pp. 352-364, 2020.

[6] P. C. Bressloff, *Waves in Neural Media,* Part of the book series: Lecture Notes on Mathematical Modelling in the Life Sciences, New York, USA, Springer, 2014.

[7] (2023) CIFAR-10 Dataset. [Online] Available: `http://www.cs.toronto.edu/~kriz/cifar.html`

[8] (2023) STL-10 Dataset. [Online] Available: `http://ai.stanford.edu/~acoates/stl10/`

[9] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, pp. 448-456, 2015.

[10] S. Ruder, "An overview of gradient descent optimization algorithms", arXiv:1609.04747, 2020.

[11] A. Al-Kababji, F. Bensaali and S. P. Dakua, "Scheduling techniques for liver segmentation: ReduceLRonPlateau vs OneCycleLR", *The 2nd International Conference on Intelligent Systems and Patterns Recognition*, Hammamet, Tunisia, pp. 204-212, 2022.

Կոնվոլյուցիոն նեյրոնային ցանցի շերտի մշակում դասակարգման առաջադրանքների արդյունավետության համար

Ռաֆայել Մ. Վեզիրյան[1] և Ռաֆայել Ն. Խաչատրյան[2]

[1]ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
[2]Questrade, Երևան, Հայաստան
e-mail: rafaelveziryan@gmail.com, raf.khachatryan4@gmail.com

Ամփոփում

Այս հոդվածը ներկայացնում է նոր 2D կոնվոլյուցիոն շերտ, որը հիմնված է նեյրոնային փոխազդեցության մասնակի դիֆերենցիալ հավասարման սկզբունքների վրա: Մեր նպատակն է օգտագործել այս շերտը՝ բարձրացնելու խորը կոնվոլյուցիոն նեյրոնային ցանցերի դասակարգման ճշգրտությունը տարբեր դասակարգման առաջադրանքների համար: Մենք հատուկ շեշտը դնում ենք ResNet ճարտարապետության մեջ դրա ինտեգրման վրա, և փորձնական գնահատումներ ենք անցկացնում CIFAR10 և STL10 տվյալների հավաքածուների վրա՝ շերտի արդյունավետությունը հաստատելու համար:

Բանալի բառեր՝ Խորը կոնվոլյուցիոն նեյրոնային ցանց, դասակարգման առաջադրանք, պատկերի մշակում, ResNet, մասնակի դիֆերենցիալ հավասարում, մալուխային հավասարում, ցանցային մեթոդ:

Разработка слоя сверточной нейронной сети для повышения эффективности задач классификации

Рафаэль М. Везирян [1] и Рафаэль Н. Хачатрян[2]

[1]Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
[2]Questrade, Ереван, Армения
e-mail: rafaelveziryan@gmail.com, raf.khachatryan4@gmail.com

Аннотация

В этой статье представлен новый двумерный сверточный слой, основанный на принципах уравнения в частных производных нейронного взаимодействия. Наша цель использовать этот слой для повышения точности классификации глубоких сверточных нейронных сетей для различных задач классификации. Мы уделяем особое внимание его интеграции в архитектуру ResNet и проводим экспериментальные оценки наборов данных CIFAR10 и STL10 для проверки его эффективности.

Ключевые слова: Глубокая сверточная нейронная сеть, задача классификации, обработка изображений, ResNet, уравнение в частных производных, уравнение кабеля, метод сетки.

Կանոններ հեղինակների համար

ՀՀ ԳԱԱ ԻԱՊԻ “Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ” պարբերականը տպագրվում է 1963 թվականից: Պարբերականում հրատարակվում են նշված ոլորտին առնչվող գիտական հոդվածներ, որոնք պարունակում են նոր՝ չհրատարակված արդյունքներ:

Հոդվածները ներկայացվում են անգլերեն՝ ձևավորված համապատասխան “ոճով” (style): Հոդվածի ձևավորման պահանջներին ավելի մանրամասն կարելի է ծանոթանալ պարբերականի կայքէջում՝ http://mpcs.sci.am/:

Rules for authors

The periodical “Mathematical Problems of Computer Science” of IIAP NAS RA has been published since 1963. Scientific articles related to the noted fields with novel and previously unpublished results are published in the periodical.

Papers should be submitted in English and prepared in the appropriate style. For more information, please visit the periodical's website at http://mpcs.sci.am/.

Правила для авторов

Журнал «Математические проблемы компьютерных наук» ИПИА НАН РА издается с 1963 года. В журнале публикуются научные статьи в указанной области, содержащие новые и ранее не опубликованные результаты.

Статьи представляются на английском языке и оформляются в соответствующем стиле. Дополнительную информацию можно получить на веб-сайте журнала: http://mpcs.sci.am/.

The electronic version of the periodical “Mathematical Problems of Computer Science” and rules for authors are available at

http://mpcs.sci.am/

Phone: (+37460) 62-35-51
Fax: (+37410) 28-20-50
E-mail: mpcs@sci.am
Website: http://mpcs.sci.am/

Ստորագրված է տպագրության՝ 23.11.2023
Թուղթը՝ օֆսեթ:
Հրատարակված է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի կողմից
Ծավալը՝ 73 էջ: Տպաքանակը՝ 100
ՀՀ ԳԱԱ ԻԱՊԻ Համակարգչային պոլիգրաֆիայի լաբորատորիա
Երևան, Պ. Սևակի 1
Հեռ. +(374 60) 623553
Գինը՝ անվճար

Подписано в печать 23.11.2023
Офсетная бумага.
Опубликовано Институтом проблем информатики и автоматизации НАН РА
Объём: 73 страниц. Тираж: 100
Лаборатория компьютерной полиграфии ИПИА НАН РА.
Ереван, П. Севака 1
Тел.: +(374 60) 623553
Цена: бесплатно

Signed in print 23.11.2023
Offset paper
Published by the Institute for Informatics and Automation Problems of NAS RA
Volume: 73 pages
Circulation: 100
Computer Printing Lab of IIAP NAS RA
Yerevan, 1, P. Sevak str.
Phone: +(374 60) 623553
Free of charge