Implementation of Deep Learning Methods in the Tasks of Ensuring Information and Psychological Safety for Operators of Automated Railway Traffic Control Systems

Konstantin O. Gnidko
MSA named after A.F.Mozhaysky
St.Petersburg, Zhdanovskaya St,13, 197198
greeny598@gmail.com

Mikhail A. Eremeev
MSA named after A.F.Mozhaysky
St.Petersburg, Zhdanovskaya St, 13, 197198
m_eremeev@mail.ru

Abstract

This paper discusses the application of deep learning methods to highlight complex patterns in the feature space of recorded parameters of the information environment and the observed psychophysiological parameters to solve the problem of identifying potentially dangerous effects on operators of automated railway traffic control systems. The results of the application of multivariate analysis to experimental data in order to highlight the most informative features and subsequently train the neural network are described. The structure of a convolutional neural network is presented, which is potentially capable, within the framework of the deep learning paradigm, to generalize at different levels of the hierarchy the poorly formalized features that distinguish harmful media content that can affect the psyche and physiological state of users.

1. Introduction

Being a powerful tool for understanding and transforming the world and man himself, information technology at the same time has become a serious threat. The ubiquity of computer networks and mass communication media has repeatedly strengthened the possibilities of remote influence on the human psyche. This led to the emergence of a new class of security threats – a harmful informational and psychological impact on the consciousness and subconscious of the personnel of automated control systems, including the field of railway communication. Being critical for the country's economy, the railway system is a very probable target for terrorist acts and other subversive activities of destructive forces interested in destabilizing the socio-economic and socio-political situation. This fact makes the task of ensuring the safety of all elements of the transport system, especially its governing bodies, extremely urgent.

Traditionally the information security is considered as security of information systems, economic and legal structures of the states, their invulnerability for negative information impact of the opponent. But behind these diverse objects the human's personality is usually lost. And the person itself has to become a subject of close attention and protection from possible threats, including information and psychological, in the era of ultimately rapid development of telecommunication systems.

Thus, not only the problem of information security, but also the problem of protection from information has recently acquired an international scale and strategic nature.

2 Threats of potentially harmful multimedia content

Based on previous research, e.g. [Gni2015] and [Ost2019] we can enumerate the subsets of classes of potentially harmful multimedia content as follows.

Texts. The most significant parameters determining the effect of the text on the psychophysiological state of a person include: phonosemantic characteristics that are not reducible to the semantic content of the text; sentiment properties of the text; fractal properties that determine the measure of the text's suggestiveness.

Video streams. The visual analyzer through which about 90% of all the processed information comes is most important for the activity of the human-machine operator. Vision allows to perceive the shape, color, brightness and movement of objects. The possibility of visual perception is determined by the energy, spatial, temporal and informational properties of signals

received by the operator. The combination of these properties and the dynamics of their changes over time (the structure of the video stream) determine the information that the visual signal transfers into the conscious and unconscious space of the operator. The incident in Japan that led to the hospitalization of more than 700 children after watching the 'Pokemon' cartoon series on December 16, 2007 is well known. It was caused by flashing of red and blue color spots for about 10 seconds which resulted in the effect of resonance with the main frequencies of the brain activity.

Experimentally proved that visual inserts as well as light frequency stimulation can affect the psycho-physiological state and subconscious of a person. In this regard, the following types of artefacts in the video stream are subject to filtering: hidden frames; part-frame incuts; brightness fluctuations (flicker) in the range of biologically sensitive frequencies.

Audiostreams. Malicious content to be filtered in audio streams includes: audio suggestion (from Latin 'suggestio') { the process and result of reproducing and perceiving special audio information, which leads to a significant decrease in the threshold of critical perception of the user and changes his emotional and psycho-physiological state; harmful binaural rhythms in the range of biologically sensitive frequencies (the so-called 'digital drugs'); hidden subliminal incuts in the audio stream.

A list of the types of potentially dangerous impacts that can be embedded in multimedia content, along with the relevant features and a summary of detection methods are presented in Table 1.

Despite the availability of particular methods for recognizing potentially harmful objects in multimedia data streams, analysis showed that currently there are no effective procedures for detecting dangerous informational and psychological states of automated control system operators by output signals available for monitoring under conditions of partial observability of the internal states of the complex 'subject – media' system, heterogeneity and fuzziness of complexly structured psychophysiological data. The presence of an anthropogenic factor moves such systems into a class of poorly formalized ones. Systems of this type function under conditions of uncertainty, characterized by a lack of information about the informative indicators of threatened states, necessary for formalizing the processes taking place in such systems. On the one hand, the uncertainty is caused by the insufficiency or complete absence of methods and means of measuring the internal variables of the state of anthropogenic elements of the system in the phase space of large dimension, and on the other hand, ignorance of the patterns of mental processes due to their complexity and stealth from the observer.

3 Deep learning in the tasks of ensuring information and psychological safety for operators of automated systems

Taking into consideration the above mentioned, we have proposed an approach that includes the use of non-invasive means of monitoring the psychophysiological parameters of an automated system operator and deep learning technology to automatically identify complex (deep) patterns that correspond to potentially dangerous states of the human psyche.

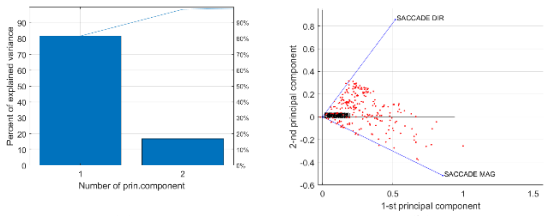
The experimental part of the study included demonstration of utterly emotionally significant visual stimuli (positive, negative and neutral) to the volunteers while recording their oculometry data.

To obtain experimental data a precise Gazepoint oculograph [Gaz2015] was used. It allowed, in particular, to record such psychophysiological indicators of subjects' reactions to the presented stimuli as directions and lengths of saccades (fast, strictly coordinated eye movements occurring simultaneously and in one direction), the coordinates of the eye fixation points on the monitor, the number and duration of blinks, the diameter of the pupils and a number of others. In total, 25 indicators were included in the apriori feature set. The number of records in the experimental database (after preliminary filtering of invalid values) exceeded 300,000. To highlight the most informative features and reduce the dimension of the feature space, the principal component analysis (PCA) was used, which allowed us to switch from the original correlated features to a strictly orthogonal basis of linearly independent factors (the main components that maximize the variance of the source data) and discard redundant, uninformative features. The result of this transformation is shown in Figure 1.

The Pareto diagram in Figure 1 (a) shows that 95% of the initial dispersion is explained by only two first principal components, and Figure 1 (b) shows the mapping of the raw experimental data into an updated basis of the first two principal components. The primary features that make the most significant contribution to

Table 1: Types of potentially dangerous impacts that can be embedded in multimedia content

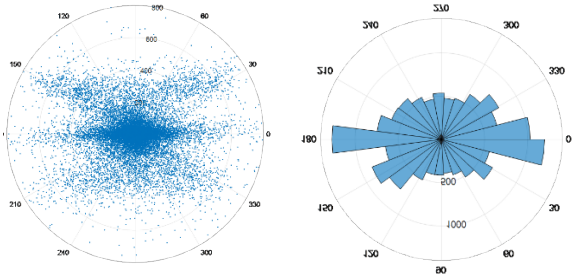
Type of impact	Indicator	Method of detection
Text		
Suggestive	Stable rhythm and self-similarity of separate text fragments at the phonetic and lexicographical levels.	Conversion of the text to the form of a series of integers with the subsequent calculation of the Hurst index by the method of normalized scope (*R/S*-analysis).
Emotional	a) Prevalence in the text of letters (sounds) with a negative phonosemantic coloring. b) The presence of a large number of emotionally strong words and markers ('emoticons', exclamation marks, etc. in the text).	a) Calculation of phonosemantic evaluation (according to A. Zhuravlev). b) Assessment of the text tonality by the method of sentiment analysis [The2017].
Laboriousness of decoding	Significant deviation of text statistical indicators from the evolution model of speech code.	Calculation of static laboriousness of decoding and dynamic laboriousness of decoding estimates in the paradigm of the evolutionary speech code.
Video		
Flicker in the range of biologically sensitive frequencies	Periodic change in the brightness of the light flux, affecting the visual analyzer, with a frequency close to the natural frequencies of the functioning of the brain	Calculation of the integral brightness of pixels for each frame of the analyzed video sequence, conversion to the amplitude-frequency representation of the integral brightness of the analyzed frame sequence through the fast Fourier transform; comparison of the obtained frequency response with the parameters of the psycho-visual model.
Hidden frames ('25-th frames')	The presence in the videostream of one or two consecutive frames with a total time of demonstration not more than 113 ms, significantly different from both previous and subsequent frames in content.	a) Calculation of the integral brightness of frame-differences and the detection of incuts on the 'Λ' and ' ΛΛ ' shaped patterns on the graph; b) detection of frame incuts based on perceptual hashing.
Part-frame incuts	Presence in the video sequence of consecutive frames differing in short-term (no more than 113 ms) demonstration of separate fragments of images (symbols, faces, etc.).	Calculation of frame-differences for all frames of a video sequence; localization of suspicious for the presence of a dissonant insert by the presence of a burst on the chart of detailed wavelet coefficients.
Audio		
Audiosuggestion	Stable rhythm and self-similarity of the audio stream at different levels of scaling.	Low-frequency filtering of the signal based on a discrete wavelet transform with the subsequent calculation of the fractal Hurst index based on the obtained approximating wavelet coefficients.
Binaural beats ('digital drugs')	Long (more than 10 seconds) transmission on the right and left stereo channels of audio streams with an absolute difference in main frequencies lying in the bio-efficient range of 0-25 Hz.	Fourier transform of the right and left stereo channels of the audio signal; calculation of the absolute difference in main frequencies; comparison of the obtained value with a range of human biologically sensitive frequencies.
Hidden subliminal incuts in the audio stream	The presence in the audio stream of imperceptible artificially inserted audio objects (commands) of short duration, discordant with a background.	Pre-filtering of an audio stream by means of discrete wavelet transform followed by recognition of anomalies in detailed wavelet coefficients.



(a) (b)

Figure 1: Pareto diagram (a) and representation of raw oculogram data in the space of two first principal components (b)

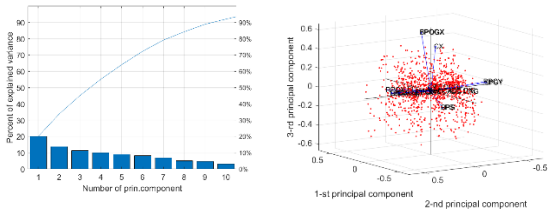
these components are also shown in Figure 1(b) – the saccade magnitudes (SACCADE MAG) and their angular directions (SACCADE DIR). For convenience of further processing, they are presented in the polar coordinate system (Figure 2 (a)), as well as in the form of a polar histogram (Figure 2(b)).



(a) (b)

Figure 2: Representation of the direction and length of saccades in polar coordinates (a) and in the form of a polar histogram (b)

Despite the complete agreement of the experimental results with the expected oculomotor activity of the subjects viewing images on a monitor screen, analysis of the data showed that based on only the first two principal components it was impossible to distinguish potentially harmful multimedia from the neutral ones, which would lead to significant difficulties in training the network classifier. One of the main reasons which led to this result was a large discrepancy in the dimensionality of the analyzed raw data features. For example, the direction of the saccade ranges from 0 to 360 degrees, while the pupil diameter varies from 1 to 2 mm. To avoid the effects of dimensionality discrepancy we performed z-normalization of the data by formula $z = \frac{x-\bar{X}}{\sigma}$, where $\bar{X}$ – the sample means by the corresponding data column, σ – the standard deviation for the same column. After the normalization the principal component method was applied again (Figure 3).



(a) (b)

Figure 3: Pareto diagram (a) and representation of raw oculogram data in the space of 3 first principal components (b) after normalization

The Pareto diagram shows that the number of principal components (and the corresponding features corresponding) in this case increased to 10 (Figure 3(a)), and the visualization of data in three-dimensional space shows a significant contribution to the initial variance of such primary parameters as saccades, coordinates of the viewpoints (BPOGX and BPOGY), diameters of the left and right pupils (LPMM and RPMM). Identified informative features in this way form a working dictionary for implementing deep learning methods.

The next task is the choice of the deep learning paradigm and model hyperparameters. This problem is also one of the weakly formalizable, does not have a strict solution, and largely depends on the experience and intuition of the researcher. Taking into account the exceptional potential variety and complexity of visual stimuli that can contain negative content, it seems advisable to use the mathematical apparatus of convolutional neural networks (CNN), which was originally developed as a formal analogue of the visual cortex of the brain and has proven itself in solving problems of image classification [San2019]. A general view and an enlarged fragment of the developed classifying convolutional neural network containing 144 layers and 168 connections is shown in Figure 4.

A distinguishing feature of the developed neural network is the use of the Leaky ReLU activation function in linear rectification units to avoid the problem of overfitting of individual layers and the neural network as a whole. The procedure for training of the resulting classifier and the results of evaluating its quality are beyond the scope of this article and will be examined in detail in our subsequent works.

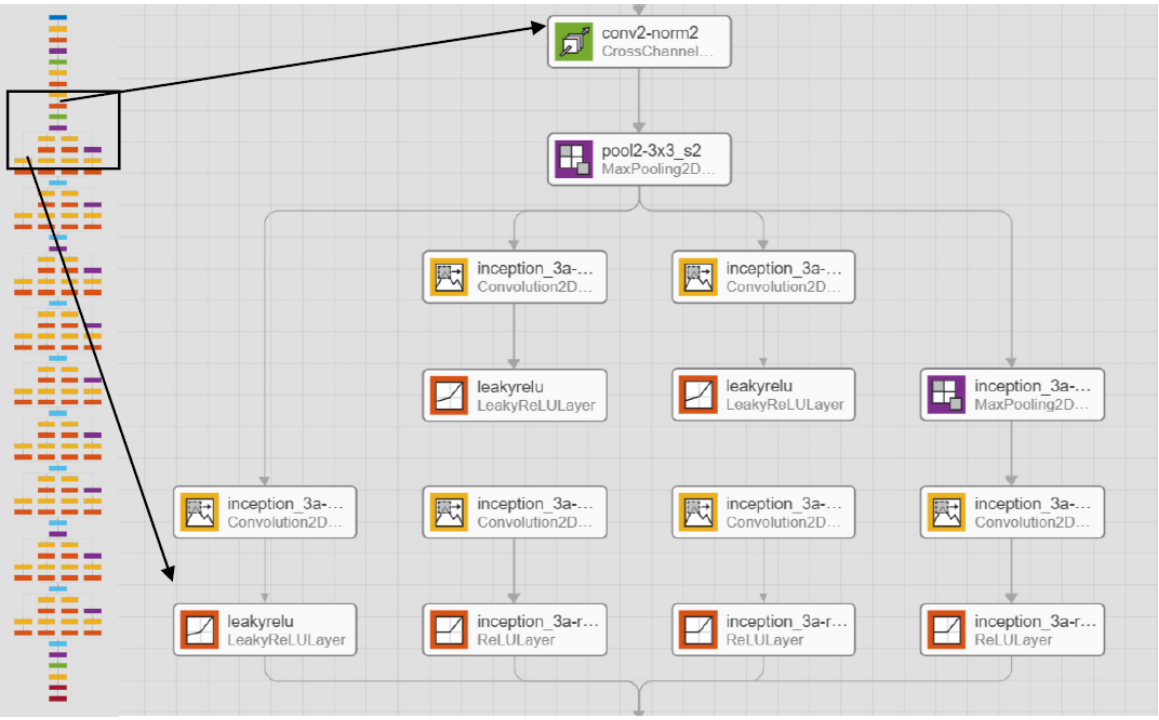


Figure 4: The structure of a convolutional neural network for the automatic classification of potentially dangerous information-psychological states

4 Summary

Thus, the work presents the results of a study devoted to the problem of protecting operators of automated railway control systems from potentially harmful information and psychological influences based on psychophysiological monitoring by means of Gazepoint software and hardware complex and deep learning methods. It has been determined that the most informative of all the parameters recorded during the experiment were the coordinates of the gaze fixation points, the pupil diameters of the subject and the characteristics of saccades. A specific model of a convolutional neural network has been proposed, which, using the listed features as input values, can be trained to detect graphic content that can harm the psyche and physiological state of automated systems operators. In order to reduce the probability of the CNN overfitting the partial replacement of the most commonly used ReLU activation function with the Leaky ReLU is provided. Testing the quality of the developed convolutional neural network is the goal of future research.

4.1 Acknowledgements

The study was carried out with the _nancial support of the Russian Foundation for Basic Research, project № 18-29-22064\18.

References

[Com79] Comer D. The ubiquitous b-tree. Computing Surveys, 11(2):121–137, June 1979.

[Gni2015] Gnidko K.O., Lomako A.G. Kontrol' potencial'no opasnogo informacionno-psihologicheskogo vozdejstvija na individual'noe i gruppovoe soznanie potrebitelej mul'timedijnogo kontenta [Control of potentially dangerous information and psychological impact on the individual and group consciousness of multimedia content consumers]. *Trudy SPIIRAN [Proceedings of SPIIRAS]*, 38:9–33, 2015 (In Russ.).

[Ost2019] Ostroushko A., Karpuhin D., Merkushova O., et al. Counteraction to the harmful information impact on the psyche of children in the Internet. *Revista ESPACIOS* 40, 2019

[Gaz2015] Eye Tracking System Technology For Everyone - UX Testing. In: Gazepoint. https://www.gazept.com/. Accessed 20 Nov 2019

[The2017] Thelwall M. The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength Cyberemotions. Springer, Pp. 119-134, 2017

[San2019] Santamaria-Pang A, Li Q, Sui Y, et al. Systems and methods for automatic generation of training sets for machine interpretation of images. *Google Patents*, 2019