Приложение 09

ABSTRACTS OF ARTICLES PRESENTED IN THE COMPETITION

Galina Dimitrova Momcheva

A candidate for the competition of academic position "Associate Professsor"
in the professional area 4.6. „Computing and Computer Science" (Neural Networks Architectures)
initiated by Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
(announced in State Gazette, 69 (80.) / 11.08.2023)

НАУЧНИ СТАТИИ

1. Fuzzy U-Net Neural Network Design for Image Segmentation

Kirichev, M., Slavov, T., Momcheva, G.

Lecture Notes in Networks and Systems, 2022, 374 LNNS, pp. 177–184

Abstract: One of the problems in biomedical image analysis is the problem of nuclei segmentation. The annotation of nuclei by hand has proven itself very time consuming with varying results depending on many factors. More recently convolutional neural networks have made the problem of automatic image segmentation easier, faster and more reliable. In this article, an extension to the standard U-Net is proposed which aims at improving the quality of biomedical image segmentation. By integrating Fuzzy computations in the standard U-Net architecture we have achieved even better accuracies than the ones reached by the base architecture. The Fuzzy Layers resemble a sense of uncertainty, which is already seen in the real world. This allows for a more precise detection and instance segmentation of cellular nuclei. When the original U-Net was first developed one of its focuses was to be trainable with a small dataset. This quality has been proved useful when undertaking the task of Biomedical Image Segmentation. The model was trained with Kaggle 2018 Dataset. The segmentation process omes right after, including the following two steps: 1) creating a prediction matrix, 2) thresholding the matrix to attain a visual result. The second step exhausts the following techniques: Manual Thresholding; Adaptive Threasholding; Gaussian Threasholding, and Otsu Thresholding. These results point out which threasholding technique, combined with a certain set of Fuzzy Layers, yelds the best possible results in tearms of accuracy of the model.

2. Fuzzy U-Net Neural Network Architecture Optimization for Image Segmentation

Kirichev, M.M., Slavov, T.S., Momcheva, G.D.

IOP Conference Series: Materials Science and Engineering, 2021, 1031(1), 012077

Abstract. In this article, the optimization of the modified U-Net neural network model extended with fuzzy layers has been studied with the usage of Grid search and Keras tuner. The article is a continuation of previous work where the model is suggested and explored. From one point of view, the research is focused on the optimization of Fuzzy Layers embedded in the U-Net model in order to find the better neural network architecture for nuclei segmentation in the research work in BioMed Varna R&D ecosystem for the segmentation of cellular nuclei. At the same time from a global perspective, this experiment is a part of the bigger one for the searching of new neural network architecture design techniques.

3. Sentiment detection with FedMD: Federated learning via model distillation

Tsankova, P., Momcheva, G.

CEUR Workshop Proceedings, 2020, 2656, pp. 236–247

Abstract. Federated learning is a distributed machine learning technique in which client devices train models locally without sharing any data, except for parameter changes, which get aggregated to a central model. This privacy-preserving approach has a huge potential for reconciling the need for large Deep Learning datasets with the increasing sensitivity of data ownership. Our paper takes the novel FedMD (Federated Learning via Model Distillation) algorithm and applies it for the first time to the fi eld of Natural Language Processing. The results are promising with regards to solving the data heterogeneity and model personalization challenges by introducing client-specifi c models and collaborative learning realized through model distillation. The resulting small gap between the FedMD results and the non-FedMD implementation is compensated by the smaller amount of training data for the FedMD models and the successful preservation of privacy for locally available data.

4. Hyperparameter adjustment in regression neural networks for predicting support case durations

Hristov, H., Momcheva, G.

AIP Conference Proceedings, 2021, 2333, 070015

Abstract. Regression is a powerful technique for predicting a single scalar value with a high degree of certainty. A regression model requires a dataset that consists of only numeric features. However, datasets frequently contain bothnumeric and categorical features. This paper sets out to study several different text encoding techniques, which can solve this problem by transforming a certain text value into a corresponding numerical value that is statistically sane in relation to the dataset. The scenario we study is a neural network regression model predicting support case durations. Many of the dataset's features are indeed categorical such as issue description, team, username, severity and impact among others. The most challenging aspect of the encoding is the resulting change of the dimensionality of the dataset. Every encoding method affects the dimensionality in various degrees depending on the feature cardinality. This result is the main challenge for the tuning of the neural network hyperparameters. We must make such a setup that can robustly handle the altered dataset. The paper compares five different approaches for encoding: one-hot, hashing, binary, target and entity embeddings. The hyperparameter settings for each approach are presented by using common neural network performance metrics and a baseline neural network setup. We can conclude that a moderately increased dimensionality can enhance the model's predictive power as observed in the case of the binary and the hashing encoder

5. Sustainability of Research-based Ecosystem

Ivanova, A., Momcheva, G., Zhekova, R., Tankova, E., Pavlov, S.

AIP Conference Proceedings, 2022, 2505, 060008

Abstract. The aim of the paper is to study sustainability of the research-based entrepreneurial ecosystem BioMed-Varna. The BioMed-Varna is focused on interdisciplinary research in Biomedical Image Analysis, Computational Life Sciences, and Neuroscience. Its main activities include multidisciplinary scientific research, educational projects, development of STEM

practices and events/cases for educational institutions, promotion of innovation and entrepreneurial culture, and support for its members' scientific advancement. This paper investigates the development and performs a sustainability appraisal of the research-based entrepreneurial ecosystem using real-world data. The analysis applied by Social network analysis (SNA), reveals the participants' roles, activities and relationships. Using Key Point Indicators (KPIs) and network-based models, this research can support decisions about necessary changes in the ecosystem. Results can be implemented for running evaluation as well as for predictive sustainability assessment. The methodology can be transferred for further research on other regional ecosystems.

6. Feasibility of Haralick's Texture Features for the Classification of Chromogenic In-situ Hybridization Images

Pavlov, S., Momcheva, G., Burlakova, P., Atanasov, S., Stoyanov, D., Ivanov, M., Tonchev, A.

Proceedings of the International Conference on Biomedical Innovations and Applications, BIA 2020, pp. 65–68, 9244282

Abstract—This paper presents a proof of concept for the usefulness of second-order texture features for the qualitative analysis and classification of chromogenic in-situ hybridization whole slide images in high-throughput imaging experiments. The challenge is that currently, the gold standard for gene expression grading in such images is expert assessment. The idea of the research team is to use different approaches in the analysis of these images that will be used for structural segmentation and functional analysis in gene expression. The article presents such perspective idea to select a number of textural features that are going to be used for classification. In our experiment, natural grouping of image samples (tiles) depending on their local texture properties was explored in an unsupervised classification procedure. The features are reduced to two dimensions with fuzzy c-means clustering. The overall conclusion of this experiment is that Haralick features are a viable choice for classification and analysis of chromogenic in-situ hybridization image data. The principal component analysis approach produced slightly more “understandable” from an annotator’s point of view classes.

7. Gabor Features for the Classification and Evaluation of Chromogenic In-Situ Hybridization Images

Pavlov, S., Momcheva, G., Burlakova, P., Atanasov, S., Stoyanov, D., Ivanov, M., Tonchev, A.

Lecture Notes in Networks and Systems, 2022, 374 LNNS, pp. 375–383

Abstract. High-throughput chromogenic in-situ hybridization (CISH) is a brightfield microscopic technique that reveals the spatial distribution of gene expression in animal cells and tissues by an easily detectable coloured precipitate. The “golden standard” for the grading of CISH-stained tissues involves qualitative scoring by a domain expert. This method is biased, suffers from low reproducibility, and lowers the efficiency of high-throughput experiments. A few quantitative image analysis approaches resolve these issues, but the proposed methods are sensitive to experimental conditions or require expert adjustment of multiple parameters.

The idea of our research team is to extract textural information from CISH-images that will be used to generate a feature space for semantic segmentation and functional analysis of gene expression. In our current work, we explore the idea by unsupervised classification based on features generated via Gabor energy filters. The tissue was divided into overlapping 150 μm tiles and processed with a Gabor filter bank (5 wavelengths, 16 directions, bandwidth 1.4). The results for the 16 directions at each wavelength were combined by a maximum superposition into a single image, and the mean grey value, standard deviation and entropy were measured. After appropriate dimensionality reduction, the tiles were classified by a fuzzy C-means algorithm. Four experts without prior knowledge of the classification results evaluated the strength and pattern of gene expression of a set of randomly selected tiles, and independently each class in the original whole-slide images. A comparison between the class-scale and tile-scale evaluations was used to assess the usefulness of the selected features.

8. Text and source readability-A step to cognition

Momcheva, G., Spasova, V., Ivanova, A., Zhelyazkov, M.

AIP Conference Proceedings, 2021, 2333, 070013

Abstract. The article aims to explore the relationship between readability of texts from instructions (assignments texts, technical documentation, law/regulation texts) and the corresponding source code (task solution) from a particular education resource for programming on the topic with neural/positive sentiment 'identity documents'. The article examines two types of studies: one of

applying software metrics and second one from empirical research of text understanding for a target group of 15 years old students (on texts and programming source code). The results are a strong foundation for continuing research in order to receive objective methodology for readability for the people of this age. This has both research and also publishing companies' business added value as objective recommendations for the authors of the textbooks and also for teachers in Computer Science and languages for general teaching practices.

9. Application of process mining approach to the developmental process of the roundworm C. elegans

Chervenkov, T., Pavlov, S., Marinov, D., Hristov, H., Momcheva, G.

Proceedings of the International Conference on Biomedical Innovations and Applications, BIA 2021, pp. 51–53

Abstract—Process mining is an analytical approach which stems from and converges on data science and process modelling. Initially incepted to support business process management, however process mining approach is universal and applicable to other fields. It was already discerned that process mining techniques share similarities with such used in bioinformatics and that the emerging process mining discipline can benefit from applying techniques developed in computational biology [1]. Herein however, we demonstrate the reverse: that process mining can be applied for the study biological processes. As process mining operates on event logs in order to analyze a particular biological process it is necessary to transform the information for a sequence of biological events into an event log. For this study we applied process mining techniques to a developmental dataset from the lineage-resolved molecular atlas of the round

worm C. elegans [2]. The single-cell temporal gene expression data was transformed into event log and analyzed with process mining tools. We show that application of process mining to biological processes is feasible, yet the presentation of the results with current tools is not suitable for the high information content of the particular biological process and this hampers further extraction of knowledge. We conclude that the application of process mining to biological processes would be beneficial for both fields.

10. Survey of Information Technology Undergraduate Degree Programs in Canada

Marinova, R., Momcheva, G.

2019 IEEE Canadian Conference of Electrical and Computer Engineering, CCECE 2019, 8861715

Abstract—A survey of Information Technology (IT) undergraduate degree programs in Canada is presented. IT Bachelor degrees have not been very popular in Canada although they have existed for a number of years in other parts of the world, including the United States (US) and the European Union (EU). This survey is looking into the current status of IT undergraduate degree programs and their curricula.

11. Vendor Cybersecurity Risk Assessment in an Autonomous Mobility Ecosystem

Tzoneva, A., Momcheva, G., Stoyanov, B.

10th International Scientific Conference on Computer Science (COMSCI), 2022, DOI: 10.1109/COMSCI55378.2022.9912588

Abstract - Vendor cybersecurity risk assessment is of critical importance to smart city infrastructure and sustainability of the autonomous mobility ecosystem. Lack of engagement in cybersecurity policies and process implementation by the tier companies providing hardware or services to OEMs within this ecosystem poses a significant risk to not only the individual companies but to the ecosystem overall. The proposed quantitative method of estimating

cybersecurity risk allows vendors to have visibility to the financial risk associated with potential threats and to consequently allocate adequate resources to cybersecurity. It facilitates faster implementation of defense measures and provides a useful tool in the vendor selection process. The paper focuses on cybersecurity risk assessment as a critical part of the overall company mission to create a sustainable structure for maintaining cybersecurity health. Compound cybersecurity risk and impact on company operations as outputs of this quantitative analysis present a unique opportunity to strategically plan and make informed decisions towards acquiring a reputable position in a sustainable ecosystem. This method provides attack trees and assigns a risk factor to each vendor thus offering a competitive advantage and an insight into the supply chain risk map. This is an innovative way to look at vendor cybersecurity posture. Through a selection of unique industry specific parameters and a modular approach, this risk assessment model can be employed as a tool to navigate the supply base and prevent significant financial cost. It generates synergies within the connected vehicle ecosystem leading to a safe and sustainable economy.

12. Voronoi Diagrams and Perlin Noise for Simulation of Irregular Artefacts in Microscope Scans

Alreni, A., Momcheva, G., Pavlov, S.

9th International Conference on Bioimaging, Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies - (Volume 2), 2022, DOI: 10.5220/0010833 000003123

Artefacts are a common occurrence in microscopic images and scans used in life science research. The artefacts may be regular and irregular and arise from different sources: distortions of the illumination field, optical aberrations, foreign particles in the illumination and optical path, errors, irregularities during the processing and staining phases, et cetera. While several computational approaches for dealing with patterned distortions exist, there is no universal, efficient, reliable, and facile method for removing irregular artefacts.

This leaves life scientists within cumbersome predicaments, wastes valuable time, and may alter the analysis results. In this article, the authors outline a systematic way to introduce synthetic

irregular artefacts in microscopic scans via Perlin Noise and Voronoi Diagrams. The reasoning behind such a task is to produce pairs of “successful” and manufactured “failed” image counterparts to be used as training pairs in an artificial neural network tuned for artefact removal. At the moment, the outlined method only works for grayscale
images.

13. Advanced Image Processing Techniques for the Detection and Monitoring of TV Datacenter System

Mohamed Fakhreddine; Teodora Bakardjieva; Galina Momcheva

TV Datacenter is the primary storage facility where servers for data storage are located. TV Datacenters have exploded in popularity over the last decade, becoming the epicenter of the technical landscape. As the size and ability of the centers grow, the complexity in managing them grows as well. The information and functionality grid are in danger of imploding, with potentially catastrophic consequences. As a result, solutions that can keep pace with the growth of data delivery, as well as Datacenter storage sizes, are needed. The main objective behind this research study is to propose a complete and advanced system for accommodating the enlargement of Datacenters, for enhancing data delivery in Datacenters, and for detection of errors and problems in TV Datacenters. Moreover, the system will also perform advanced and robust monitoring of TV Datacenters. The significance of the study relies on the algorithms that are proposed since these algorithms are completely novel as well as they are highly performant. In addition, these algorithms incorporate new technologies of image processing. The results obtained using the algorithms showed a detection rate obtained of 83.33% which can be considered to be an indicator of the high performance of the proposed detection system. Moreover, according to the results, the minimum time of replacement is zero seconds meaning that the replacement was done automatically after the detection of the corruption in the video. The results also showed that for all videos, the detected number of frames were all successfully replaced. The frame replacement showed 100% efficiency for every video.

14. Multi-Activation Dendritic Neural Network (MA-DNN) Working Example of Dendritic-Based Artificial Neural Network

Konstantin Tomov, Galina Momcheva

Abstract: Throughout the years neural networks have been based on the perceptron model of the artificial neuron. Attempts to stray from it are few to none. The perceptron simply works and that has discouraged research around other neuron models. New discoveries highlight the importance of dendrites in the neuron, but the perceptron model does not include them. This brings us to the goal of the paper which is to present and test different models of artificial neurons that utilize dendrites to create an artificial neuron that better represents the biological neuron. The authors propose two models. One is made with the purpose of testing the idea of the dendritic neuron. The distinguishing feature of the second model is that it implements activation functions after its dendrites. Results from the second model suggest that it performs as well as or even better than the perceptron model.

15. Utilizing Autoencoders for Analysis and Classification of Microscopic In Situ Hybridization Images

Aleksandar A. Yanev, Galina D. Momcheva, Stoyan P. Pavlov

Abstract: Currently, analysis of microscopic In Situ Hybridization (ISH) images is done manually by experts. Precise evaluation and classification of such microscopic images can ease experts' work and reveal further insights about the data. In this work, we propose a deep-learning workflow to detect and classify areas of microscopic images with similar levels of gene expression. Analysis of the data is done by employing a type of ANN – Deep Learning Autoencoders – suitable for unsupervised learning. The model's performance is optimised by balancing the latent layers' length and complexity and fine-tuning hyperparameters. The results are validated by adapting the mean-squared error (MSE) metric and comparison to expert's evaluation. Reconstruction of the whole-scale microscopic images is used to summarise and visualise the results.