Abstract Book

6th International Workshop on Data Science

Zagreb, Croatia, November 24, 2021

Organiser

Centre of Research Excellence for Data Science and Cooperative Systems
Research Unit for Data Science

Member Institutions

University of Zagreb Faculty of Electrical Engineering and Computing

Ruđer Bošković Institute

University of Zagreb Faculty of Science

University of Zagreb Faculty of Transport and Traffic Sciences

Catholic University of Croatia

University of Split Faculty of Electrical Engineering, Mechanical Engineering and Naval Architeture

University of Rijeka Centre for Advanced Computing and Modelling

University of Rijeka University of Rijeka Faculty of Civil Engineering

Josip Juraj Strossmayer University of Osijek Faculty of Electrical Engineering, Computer Science and Information Technology

Sponsors

Ministry of Science and Education, Republic of Croatia
University of Zagreb Faculty of Electrical Engineering and Computing, Croatia

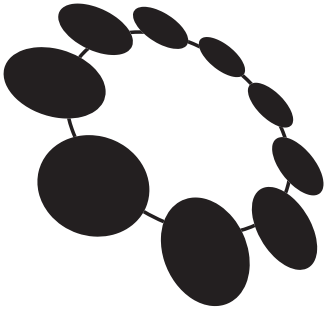
Editors

Sven Lončarić (sven.loncaric@fer.hr)
University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3, HR-10000 Zagreb, Croatia

Tomislav Šmuc (tomislav.smuc@irb.hr)
Ruđer Bošković Institute
Bijenička cesta 54, HR-10000 Zagreb, Croatia

Publisher

Centre of Research Excellence for Data Science and Cooperative Systems
Research Unit for Data Science
Croatia



RESEARCH UNIT
FOR DATA SCIENCE
Centre of Research Excellence for
Data Science and Cooperative Systems

Copyright © 2021 by the Centre of Research Excellence for Data Science and Cooperative Systems. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise without the prior written permission of the Publisher.

Welcome Message

On behalf of the Organizing Committee it is with great pleasure that we extend a warm welcome to all participants of the 6th International Workshop on Data Science (IWDS 2021) in Zagreb. The workshop is organized by the Research Unit for Data Science (RUDS), Centre of Research Excellence for Data Science and Cooperative Systems. University of Zagreb Faculty of Electrical Engineering and Computing is the host organization of the IWDS 2021.

Data Science addresses the problem of knowledge extraction from structured and unstructured data, including very large data sets referred to as Big Data. Data Science is characterized by diverse applications in many disciplines. Due to worldwide industry and government need for Data Science expertise, the market demand for data scientists has soared in recent years. RUDS addresses this pressing issue by bringing together nine high-education and research institutions and the best researchers in Croatia involved in research on foundations and multidisciplinary applications of Data Science to advance the state of the art in theory, technology, and systems. RUDS mission is to become the regional leader in research and applications of Data Science to improve the quality of life and support the economic growth in Croatia. To this end, RUDS will build partnerships with academic, government, and business partners in the areas of expertise covered by the RUDS research team, such as machine learning, data mining, complex networks and social networks, bioinformatics, natural language processing, text mining, business analytics, high-performance computing, signal and image processing, and financial applications. The final beneficiaries are the scientific community, the industry, and the government. The overall objectives of the RUDS are: (1) research excellence; (2) strengthening the transfer of technology; (3) providing the industry and the government with an access to advanced computing facilities and expertise; and (4) education and training for young researchers in Data Science disciplines.

The scientific program of the workshop consists of a plenary lecture and of oral and poster presentations. The aim of the Workshop is to foster interaction of researchers and exchange of new ideas in the diverse area of data science. Furthermore, interaction and networking between researchers and professionals in industry and government organizations is encouraged. The workshop audience will include doctoral students, postdoctoral researchers, and professionals working in academia, industry, and government, who are active in the area of data

science theory and applications.

We would like to thank all authors for their contributions to the program, members of the IWDS 2021 Organizing Committee who invested their valuable time to make this event successful. We are confident that the event will truly be fruitful and memorable for everyone and we look forward to meet you in Zagreb for IWDS 2021.

Zagreb, November 2021

Sven Lončarić, Ph.D.
General Co-Chair and Director
Centre of Research Excellence for Data Science and Cooperative Systems

Tomislav Šmuc, Ph.D
General Co-Chair and Deputy Director
Research Unit for Data Science
Centre of Research Excellence for Data Science and Cooperative Systems

Contents

Organising Committee

General Co-Chairs

Sven Lončarić, University of Zagreb, Croatia
Tomislav Šmuc, Ruđer Bošković Institute, Croatia

Program Committe Chair

Marko Subašić, University of Zagreb, Croatia

Program Committee

Ana Babić, Croatia
Bojan Basrak, Croatia
Vuko Brigljević, Croatia
Zlatan Car, Croatia
Tonči Carić, Croatia
Bojana Dalbelo Bašić, Croatia
Davor Davidović, Croatia
Mirjana Domazet Lošo, Croatia
Tomislav Domazet Lošo, Croatia
Ante Đerek, Croatia
Neven Elezović, Croatia
Joao Gama, Portugal
Nikola Godinović, Croatia
Hrvoje Gold, Croatia
Sonja Grgić, Croatia
Edouard Ivanjko, Croatia
Zoran Kalafatić, Croatia
Wolfgang Ketter, The Netherlands
Mladen Kolar, USA
Ivica Kopriva, Croatia
Nada Lavrač, Slovenia
Damir Lelas, Croatia
Robert Manger, Croatia
Goran Martinović, Croatia
Branka Medved Rogina, Croatia
Igor Mekterović, Croatia
Marie-Francine Moens, Belgium
Niranjan Nagarajan, Singapore
Davor Petrinović, Croatia
Boris Podobnik, Croatia
Tomislav Pribanić, Croatia
Krešimir Pripužić, Croatia
Ivica Puljak, Croatia
Strahil Ristov, Croatia
Damir Seršić, Croatia
Karolj Skala, Croatia
Vernesa Smolčić, Croatia
Siniša Srbljić, Croatia
Marko Subašić, Croatia
Mile Šikić, Croatia
Jan Šnajder, Croatia
Hrvoje Štefančić, Croatia
Kristian Vlahoviček, Croatia
Mladen Vouk, USA
Boris Vrdoljak, Croatia
Mladen Vučić, Croatia
Vinko Zlatić, Croatia

Registration Chair	Damir Pintar, Croatia
Publications Chair	Denis Milošević, Croatia
Industry Liaison	Ana Sović, Croatia
Secretariat	Marina Rajsman, Croatia

Plenary Lecture

How NC State Contributes to Data Economy

M. Vouk

North Carolina State University, USA

Data Economy is a simple term for a very complex and very important part of today's world economy. One of many components in that echo system is data literacy and the availability of workforce trained in advanced data science and analytics. NC State has a long history in the domain of advanced analytics and data science. For example, SAS Inst. was started at NC State in mid 1970-ies; the first North American Master of Science in Analytics (MSA) was started by NC State in 2007; NC State Statistics and Computer Science departments have been at the forefront of advanced research in artificial intelligence, machine learning, and data science for a long time, etc. In 2014 NC State put in place a formal initiative in data science intended to institutionalize teaching and training in that space, and this year initiative has transformed into Data Science Academy (DSA) - a very interdisciplinary envelope for data science outreach, teaching, and research. This talk will provide an overview of DSA and of the Advanced Analytics and Data Science pathways being built by NC State. Of course, in practice, there are many jobs that require data science and advanced analytics skills at some level but may be advertised and filled with different job titles. In general, many require skills that again are probably MS or PhD level, and/or require more than entry-level experience. But, many in domain-specific areas may require only moderate data science skills. That broadens employment opportunities. In general, basic data literacy is essential for all in today's economy; moderate data literacy is probably needed for many if not most of the desk jobs we have today. When it comes to high-end data science jobs, it is a buyer's market – there is a shortage of workforce with data science skills needed. The talk will discuss some of the related parameters such as capability and capacity of workforce training pathways.

Poster Session

Multimodal Data Processing and Information Management

SmartUTX - Modules for Assisted UT Analysis

F. Milković, B. Filipović, D. Medak, L. Posilović, T. Petković, M. Subašić, M. Budimir, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, and INETEC- Institute for Nuclear Technology d.o.o., Croatia

Nondestructive testing (NDT) is a widely used approach for material and product inspection because it allows the evaluation of objects without causing any damage to them. This is the key advantage of using NDT - an object can still be used after inspection and in some cases even during inspection. This is a requirement in many cases of quality and safety control. One of the most popular types of NDT is ultrasonic testing (UT). Because of its reliability, ease of use, and safety, UT has a wide range of applications. In this case, it is used for material inspection in nuclear power plants. A drawback of UT is the large amount of data generated during testing. Expert analysts often spend weeks analyzing the data collected during inspection. Having 2D UT scans enables framing the problem as image analysis and using deep learning (DL) algorithms for computer vision (CV) significantly reduces time spent on data analysis while successfully finding all anomalies (e.g. voids and cracks) in a block of material. Several different methods and algorithms are used, based on the type of UT scans and the specific goal of the analysis. The first type of UT scans we used were B-scans which show a cross-section of the material, thereby representing the scanning results in an intuitive way and allowing the analysts simple localization of the defects. In these scans, defects usually appear as elongated shapes that are displayed with a darker color than the neighboring pixels. The amount of noise and geometry signals can sometimes make the process of flaw detection difficult but the main problem remains the amount of data that needs to be analyzed. Recently, deep learning object detectors took over this field of computer vision and are showing great results in many popular object detection challenges like COCO and PASCAL VOC. Using object detection algorithms is a concise and reliable way to assist the analysts with defect detection in B-scans and that is why employed

three popular algorithms to localize the defects from B-scans. More specifically we used one-stage detectors YOLOv3, RetinaNet, and EfficientDet-D0 which respectively achieved 80.6%, 85.5%, and 89.6% of mean average precision on our in-house UT dataset. These results were achieved using custom anchors to overcome extreme aspect ratios of defects, which gave up to 6% mAP improvement compared to using default anchors [1]. C-scans are another type of UT scan, which can be utilized to significantly optimize the process of UT analysis. C-scans consist of selected values taken from several neighboring B-scans and are meant to show the top-down perspective of the material in 2D. Therefore, it is relatively easy to quickly detect B-scans that contain at least one defect using a convolutional neural network (CNN) classifier which has been specifically made for this purpose. This way the object detection algorithms don't have to analyze all of the collected B-scans, which significantly accelerates the assisted UT analysis. Data for training the custom CNN classification model was generated with a Generative adversarial network (GAN). We report a 94.1% F1 score [2]. Most of the collected scans show a normal inner structure because defects are actually anomalies, meaning they are truly rare. To make use of all this seemingly useless data, a variational autoencoder (VAE) is used. VAE, as a type of generative DL model, learns the distributions that describe the data. In our research [3] we used an additional encoder placed on top of the VAE's decoder to encode reconstructions of the VAE. This encoder is trained to produce the same encodings as the encoder within the VAE fed with original images. If this model is trained using only normal data, then during inference anomalous data should result in poorer reconstructions and larger distances between original and reconstruction encodings. This way all the potentially anomalous B scans can be picked out and an analyst can save a vast amount of time by focusing only on those B scans. The lack of defects in real ultrasonic inspections is also a problem we addressed because it makes training new human experts on analyzing UT data a hard task. They are often trained on artificial blocks, but the production of such blocks is expensive and complex. Therefore it would be useful to be able to generate synthetic UT images using deep learning. In recent years generative networks have attracted a lot of researchers' attention. GANs have already been proven to generate realistic images in many fields, but in our research, their applicability has also been shown in the NDT field. Synthetic data of sufficient quality could also be used to improve the performance of deep defect detectors. All the aforementioned DL models were developed in Python, either using PyTorch framework or Tensorflow framework with Keras library. The models are integrated into the SignyOne application as separate modules for assisted UT analysis.

References:

1 D. Medak, L. Posilović, M. Subašić, M. Budimir and S. Lončarić, "Automated Defect Detection From Ultrasonic Images Using Deep Learning,"

in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 68, no. 10, pp. 3126-3134, Oct. 2021, doi: 10.1109/TUFFC.2021.3081750.

2 B. Filipović, F. Milković, M. Subašić, S. Lončarić, T. Petković and M. Budimir, "Automated Ultrasonic Testing of Materials based on C-scan Flaw Classification," 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 230-234, doi: 10.1109/ISPA52656.2021.9552056.

3 F. Milković, B. Filipović, M. Subašić, T. Petković, S. Lončarić and M. Budimir, "Ultrasound Anomaly Detection Based on Variational Autoencoders," 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), 2021, pp. 225-229, doi: 10.1109/ISPA52656.2021.9552041.

Finding Hamiltonian cycles with graph neural networks

F. Bosnić, M. Šikić

*Faculty of Electrical Engineering and Computing, University of Zagreb, and Genome Institute of Singapore, A*STAR, Singapore*

As deep neural networks proved successful in solving variety of task, it is tempting to test them on solving hard combinatorial problems. Research in this direction has up to now focused mostly on the 2-dimensional traveling salesman problem, but there are many similar problems, such as the Hamiltonian cycle problem or the clique problem, on which such methods would be desirable. These problems are mostly NP-hard and it would be unreasonable to expect neural networks, which are of polynomial complexity, to produce exact solutions.

However, from practical perspective it often suffices to find an approximate solution of a particular subproblem. For example, the process of genome assembly, in which a DNA molecule is to be assembles from numerous small, overlapping fragments, essentially reduces to the Hamiltonian cycle problem. But one is interested in solving this problem only for the family of graphs that can be produced during sequencing of DNAs. Naturally, such graphs have certain structural and statistical properties which could aid in solving of the problem. In addition, finding cycles covering say 90% of the graph would also be considered a success. While there are several heuristic algorithms for solving the general Hamiltonian cycle problem, better heuristic can surely be designed specifically for solving the problem of DNA assembly. Many such heuristics are used in state of the art assembly programs [1], but creating them requires great amount of effort and a profound understanding of the problem. With the help of machine learning on the other hand, these heuristics could be *trained* instead.

The assembly problem is indeed a motivation for our work, but we present a study of the Hamiltonian cycle problem on a special class of random graphs.

We give two neural network solvers for this problem based on message passing [2] and graph attention [3] neural networks. Both of these are special kinds of graph neural networks which we believe are key for handling the Hamiltonian cycle problem. Graphs are processed in an auto-regressive manner and neural networks output the next node in the cycle at every iteration. Note also that the fundamental difficulty in the Hamiltonian cycle problem comes from the topological structure of the graph (i.e. its edge structure), as opposed to the traveling salesman problem in 2 dimensions (cities are represented as coordinates in the plane and one is allowed to travel between each pair of cities) where the difficulty comes from the metric structure of Euclidean plane. One can think of these two as complementary subproblems of the general traveling salesman problem.

To examine generalization properties of our neural network models we devise a specific class of random graphs based on theoretical results regarding Erdős-Rény graphs. When sampling a graph G of size n from this class,

$$\mathbb{P}(G \text{ has at least one Hamiltonian cycle}) \xrightarrow{n} c$$

for some fixed constant $0 < c < 1$. In fact, experimental results show asymptotics to be quite stable for $n > 20$. Therefore, we expect the Hamiltonian cycle problem to be of the *same difficulty* no matter the size of graph G. This gives the model a fair chance of generalizing to larger sizes.

References:

1 R. Vaser and M. Šikic. Time- and memory-efficient genome assembly with raven. Nature Computational Science, 1(5):332–336, May 2021.

2 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In D. Precup and Y. W. Teh, editors, Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pages 1263–1272. PMLR, 06–11 Aug 2017

3 P. Velickovič, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks, 2018.

Global Design of Linear Antenna Arrays With Constrained Dynamic Range Ratio

K. Vodvarka, M. Bellotti, Jurišić, M. Vučić

University of Zagreb Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing, Unska 3, 10000 Zagreb, Croatia

Abstract: The design of antenna arrays with constrained dynamic range ratio (DRR) of excitation coefficients simplifies the array's feeding network and reduces

mutual coupling between antenna elements. However, constraining the DRR pushes small coefficients to larger values, thus deteriorating the array pattern. The deterioration can be reduced if some coefficients are allowed to take zero values, resulting in a sparse design. This work considers both of these approaches. Furthermore, it presents a global optimization of nonsparse and sparse linear antenna arrays with constrained dynamic range ratio of excitation coefficients.

Introduction: There are several approaches which keep the implementation of antenna arrays simple (see [1] and the references therein). One of them optimizes the sidelobe level while constraining the dynamic range ratio (DRR) of excitation coefficients [2]. The method referred to supports positive and negative coefficients, as well as the DRR equal to 1. However, forcing a low DRR pushes small antenna coefficients to larger values. This indicates that an improvement is possible if some coefficients are allowed to take zero values [3]. In this work we present both approaches. Introducing the DRR constraints to antenna array design leads to nonconvex optimization, which is difficult to solve globally. Here, we present a branch and bound method that utilizes efficient pruning to find optimal antenna elements with specified DRR.

Branch and bound method for design of linear antenna array with constrained DRR: Common design of linear antenna arrays utilizes the minimization of maximum sidelobe level. Such a design can be expressed as a convex optimization problem. Unfortunately, introducing DRR constraints makes the problem nonconvex and difficult to solve globally. However, if all coefficient signs are known in advance, this problem can still be expressed in a convex form. In this work, we utilize branch and bound method that systematically goes through all possible combinations of coefficient signs, which form a tree. The method is supported by efficient pruning, which results in global arrays with specified DRR in a short time. For very strict requirements set on the DRR and the sidelobe level, the proposed method returns negative antenna coefficients together with positive ones.

Constraining DRR pushes small coefficients to larger values, thus degrading the array pattern. This problem is resolved by allowing some coefficients to take zero values what leads to a sparse design. Introducing sparsity to the proposed method results in smaller sidelobe level and fewer number of nonzero antenna elements.

The presented approach is convenient when coefficients are calculated for hardware at hand which can realize a specified DRR and which allows disabling of antenna elements.

Acknowledgement: This work was supported in part by Croatian Science Foundation under the project IP-2019-04-4189 - Efficient Signal Processing Sys-

tems for Software Defined Radio and in part by the European Regional Development Fund under Grant KK.01.1.1.01.0009 (DATACROSS).

References:

1 X. Fan, J. Liang, Y. Zhang, H. C. So, and X. Zhao, "Shaped power pattern synthesis with minimization of dynamic range ratio," IEEE Trans. Antennas Propag., vol. 67, no. 5, pp. 3067–3078, May 2019.

2 M. Jurisic Bellotti and M. Vucic, "Global Optimization of Pencil Beams With Constrained Dynamic Range Ratio," in Proceedings of European Conference on Antennas and Propagation, Copenhagen, Denmark, 15-20 March 2020.

3 M. Jurisic Bellotti and M. Vucic, "Global Optimization of Sparse Pencil Beams with Constrained Dynamic Range Ratio," in Proceedings of International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 5-10 July 2020.

HD-RACE: Spray-based Local Tone Mapping operator

K. Koščević, V. Stipetić, E. Provenzi, N. Banić, M. Subašić, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, and Institute de Mathematics, Université de Bordeaux, Talence, France, and Gideon Brothers, Zagreb, Croatia

High Dynamic Range (HDR) imaging is becoming a more prominent way of storing image data. It enables the capture of a much greater range of radiance values than conventional Low Dynamic Range (LDR) images, and, therefore, provides a more accurate representation of the real world. However, since devices such as displays and printers are not all capable of displaying values that span such wide ranges of magnitudes, it is necessary to reduce the dynamic range of HDR images to properly visualize them on LDR display devices. The procedure referred to as Tone Mapping (TM) is used for such conversion, and a method that performs TM is called Tone Mapping Operator (TMO). TMOs can either apply the same mapping function to all image pixels (global TMOs) or spatially variant mapping functions based on the pixel neighborhood (local TMOs).

In this paper, a local TMO combining multiple spray-based TMOs is proposed. The proposed TMO uses the luminance values of the input HDR image and performs three types of TM with each type producing the optimal result for a specific part of the dynamic range. The propsoed TMOs are based on Random Spray Retinex (RSR) [1], Automatic Color Equalization (ACE) [2], and Naka-Rushton (NR) equation [3]. With this approach, a sufficiently high level of detail can be retained in both very dark and very bright parts of an HDR image. The proposed TMO has three parameters that control the final look of the LDR image,

which are highly dependent on the image content. Therefore, the parameter tuning is performed by the user to match her/his own preference. Additionally, a local formulation of the Naka-Rushton equation based on sprays is proposed. This ensures that all TM calculations of the proposed TMO are entirely performed on local regions represented as sprays, which are introduced in RSR.

To adapt NR for local processing semi-saturation level should not be a single fixed value but should change depending on the spray position and pixels contained within the spray. Therefore, semi-saturation level in the proposed Local NR equation is the geometric average of the arithmetic and geometric averages in each spray.

An improvement of both RSR and NR as TMOs can be achieved by the fusion of the two together. In the simplest form, this fusion can be achieved by first applying NR on an HDR image to compress the range of radiance values and then applying RSR. However, such an approach cannot be considered locally adaptive since the compression performed by the NR equation is based on the arithmetic and geometric averages of the whole image. Therefore, in this paper, local NR is used to perform the fusion and model the local adaptability. Given a spray, first local NR is applied to all its pixels, and then RSR is applied. That fusion is further denoted as NR-RSR.

An illuminant-independent slope function was proposed to improve ACE. The proposed slope function will be big in areas covered by the spray in which the variation between the minimal and the maximal intensities is large. This will allow a very efficient detail rendition of this kind of area, which is well-known to be particularly difficult to tone-map. For the given definition of the slope function and the slope, the image range must be contained in $[0, 1]$. Therefore, instead of applying ACE on the pure values of, just like in NR-RSR, local NR is first applied on the spray (denoted as NR-ACE in furhter text).

RSR, NR-RSR, and NR-ACE produce a different-looking LDR version of the same image with the following distinct features: RSR can work well in very bright parts, NR-ACE covers very dark parts, and NR-RSR works as an intermediate between the two. When combined, outputs of these TMOs can give very pleasant LDR images in both color and detail. In this paper, they are combined locally in a pixel-wise convex combination by utilizing the random sprays technique.

All of the TM operations are performed on the luminance values of an image, which is normalized to have the minimum in zero. Normalized luminance is tone mapped in three different ways by applying RSR, NR-RSR, and NR-ACE. To avoid having the pixels of a spray fall outside the bounds of an image when extracting the sprays, images are expanded by mirroring them in the vertical, horizontal, and diagonal direction. The final tone mapped luminance is computed

in two steps. Frist, the output of RSR and NR-RSR are fused in a convex combination defined by per-pixel coefficient β. Second, NR-ACE is fused with the previous output in a convex combination defined by per-pixel coefficient γ, which yields the final tone mapped luminance. The next step is to extend obtained tone mapped luminance to color image as follows

$$L_c(x) = L(x)\frac{I_c(x)}{L(x)I_c(x) + (1 - L(x))\lambda(x)}. \quad (1)$$

Here, for each color channel $c \in \{R, G, B\}$, the denominator is the convex combination between the intensity $I_c(x)$ of some pixel x in that channel and the luminance of the same pixel $\lambda(x)$ based on the $L(x)$. $L(x)$ is the tone mapped luminance having the values in range $[0, 1]$. The proposed conversion ensures that the following properties are satisfied: when $L(x) = 0$ it is true that $L_c(x) = 0$; when $L(x) = 1$ it is true that $L_c(x) = 1$; and when $L(x) = \frac{1}{2}$ pixel intensity in a color channel and pixel luminance are equally valued, i.e., that $L_c(x) = I_c(x)/(I_c(x) + \lambda(x))$.

The coefficients β and γ tune the contribution of each version of the tone-mapped input image to the final result. Both coefficients are local and derived from the luminance of each image pixel. They are computed as Gaussians with respect to the zero-minimum log-luminance of each image pixel. Both coefficients are image dependent-parameters and can be adjusted according to user preference. β is defined to be equal to 1 for pixels corresponding to the brightest image parts, and for γ it is the opposite, i.e. γ is equal to 1 in the darkest image parts. This corresponds to the observations that RSR produces the most pleasing colors in the brightest image parts, and the observations that NR-ACE produces the most pleasing colors in the darkest image parts. NR-RSR is used to tune in the intermediate parts.

Guided image filtering (GIF) is further applied to coefficients β and γ. It ensures that β and γ are smooth in the uniform regions and preserves the changes on the edges. Consequently, the transitions between fused operators are more prominent which makes tiny details in an image more vivid and removes outliers in the regions where one operator should be dominant.

Experimental results show that the proposed TMO produces pleasant results by combining key features from the three base operators. However, it is limited by the presence of several parameters and the optimal parameters change drastically with the image content. Therefore, in the current version of the proposed TMO, the optimal set of parameters is not given and a mechanism for automatic determination of the GIF window size and coefficients β and γ is yet to be explored.

The second limitation of the proposed TMO in its current form is the execution time. The majority of computations in the proposed TMO are related to spray computations for each pixel, which makes the TMO slow to execute. However, since the proposed TMO is based on the RSR, the number of computations can be reduced by incorporating existing improvements of that methods.

References:

1 E. Provenzi, M. Fierro, A. Rizzi, L. De Carli, D. Gadia, and D. Marini, "Random spray Retinex: a new Retinex implementation to investigate the local properties of the model", IEEE Transactions on image processing, vol. 16, no. 1, pp. 162-171, 2006.

2 A. Rizzi, C. Gatta, and D. Marini, "A new algorithm for unsupervised global and local color correction", Patten Recognition Letters, vol. 24, no. 11, pp. 1663-1667, 2003.

3 R. Shapley and C. Enroth-Cugell, "Visual adaptation in retinal gain controls", Progress in retinal research, vol. 3, pp. 263-346, 1984.

Fast reconstruction for PET scanner with incomplete sector set

T. Matulić, R. Bagarić, D. Seršić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, and Ruđer Bošković Institute, Zagreb, Croatia

Positron emission tomography (PET) is a noninvasive imaging modality. It is based on electron-positron annihilation. Radioactive tracer is injected in a subject. Due to electron-positron annihilation, a two high-energy photon pair (511 keV) are emitted. They are traveling along the same line, but in opposite directions. If detectors detect two photons in a short time window, the detection is valid.

PET imaging is represented by Radon transform. Analytical algorithms (backprojection [1],[2], filtered backprojection (FBP) [1], backprojection-filtration (BPF) [1], 3D reprojection (3DRP)) are faster and less demanding than iterative algorithms (maximum likelihood expectation maximization (ML-EM), ordered subset expectation maximization (OSEM)). Generally, iterative methods are mainly used for image reconstruction in PET due to stochastic nature of electron-positron annihilation [3].

For PET data acquisition, we use Raytest ClearPET scanner. Raytest ClearPET scanner is a small animal PET scanner. It total of 20 sectors can be found in mentioned PET scanner. Each sector is divided in four modules that are spread in axial direction and each module has two layers of photon detectors (crystals)

in radial direction. In each layer is 8×8 matrix of photon detectors. Additionally, every even sector is shifted in axial direction by half the length of a single module which linearizes the axial sensitivity of the scanner. A total of 48 rings of detectors are possible. The Raytest ClearPET scanner which we use to measure and collect data is only partially equipped. There are only 8 sectors instead of maximum 20, four on each side of the ring.

Significant distortion can be seen in the reconstructed image when using (filtered) backprojection algorithm with only 8 sectors active. To cancel the unwanted effects, a compensation must be done. Our goal is to generate a compensation surface (a white surface) - an image that is obtained if the distribution of radioactive material is constant throughout the entire scanning region.

For each pair of crystals (i, j), we assign the probability density function (PDF) $p_{i,j}(r, \phi)$. PDF describes the probability of the position of a point source while the detection is done with pair of crystals (i, j). ClearPET scanner is rotating in tangential direction and, therefore, must be taken into account when producing a compensation surface. Therefore, we need to calculate $p(r) = C \int_0^{2\pi} \sum_i \sum_j p_{i,j}(r, \phi) d\phi$ where $p(r)$ describes the PDF of the position of a point source at a distance of r from the origin. The summation is done over all allowed combinations of pairs of crystals. The constant C is the normalization constant of the PDF.

The PDF $p_{i,j}(r, \phi)$ can be approximated by a simulation. Dithering was applied on simulated data due to physical size of each crystal. For each LOR we produced 10 dithered LORs. This is done since a point source can be detected by the same pair of crystals but in a different physical location. Each dithered LOR is discretized onto a rectangular grid (512×512 pixels). We formed a 2D histogram $\tilde{p}(x, y)$ by counting LORs. Histogram image $\tilde{p}(x, y) = \tilde{p}(r)$ has central symmetry due to integration with respect to ϕ. Furthermore, the rectangular pixel grid was transformed to polar coordinates. Since the histogram has central symmetry, we ignore angle ϕ and use only radius r for the polynomial fitting. The polynomial regression model was used and it has been experimentally established that a polynomial of 24th degree is the best fit. This part of algorithm needs to be done only once for given ClearPET scanner and it procedures a compensation surface.

Similarly, on the measured data we applied dithering to minimize the quantization error. Dithering was applied due to physical size of each crystal and uncertainty of actual location of LORs. For each LOR we produced 20 dithered LORs. Afterward, backprojection is done by discretization of each (dithered) LOR onto a rectangular grid. Similarly, the obtained image $I(x, y)$ can be interpreted as the non-normalized PDF of the occurrence of a point source at (x, y)

for the measurement.

The compensation is done as $I_C(x,y) = \frac{I(x,y)}{p(x,y)}$. In this way, we amplified areas with a smaller number of LORs and attenuated areas with a larger number of LORs. The reason for different LOR count in different part of the PET scanner is reduced number of sector. By analogy with BPF, a ramp filter was applied to a compensated image. To reduce unwanted noise we chose Gaussian filte. Finally, to enhance the compensated image a soft thresholding was applied.

The entire reconstruction algorithm for all 48 possible rings of detectors took 3.17 seconds on Threadripper 3960x with 128GB of RAM and 2 × GeForce RTX 2080 Ti with 11GB of RAM. CUDA was used for acceleration of the algorithm.

References:

1 D. Panetta and N. Camarlinghi, "3D Image Reconstruction for CT and PET: A Practical Guide with Python," CRC Press, pp. 29-51, pp. 78-82, 2020.

2 M. Defrise and P. Kinahan, "Data Acquisition and Image Reconstruction for 3D PET," in B. Bendriem and D.W. Townsend, "The Theory and Practice of 3D PET,", Developments in Nuclear Medicine, vol 32. Springer, 1998.

3 Y. Wang et al. "An improved PET image reconstruction method based on super-resolution," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 946, 2019.

Diesel spray segmentation with cone angle determination

F. Huzjan, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Although the development of internal combustion engines for passenger cars is significantly reduced due to electrification, their development is still essential for the heavy-duty transport sector. Spray systems and injection strategies contribute to the overall engine efficiency, combustion process, and pollutant formation inside the internal combustion engines [1]. For optical measurement, different techniques are being implemented, such as shadowgraphs, Schlieren photography, scattering suppression, laser-induced fluorescence, ballistic imaging, and x-ray imaging techniques [2]. After the images are captured with high-speed imaging, image processing algorithms are applied in order to define spray parameters such as spray cone angle and spray penetration. The dataset consists of 200 images that were collected during one spray injection. The images have a resolution of 768x512 pixels and are in RGB format. The segmentation masks were labeled by domain experts. An image is segmented with three methods: K-Means clustering, Otsu method, and max entropy segmentation. Two new methods for cone

angle determination are proposed and compared with two currently published methods for cone angle determination. Proposed methods consist of edge detection and line fitting over a segmented image [2]. After the image is segmented, it is preprocessed with filters and rotation to obtain the right-oriented image. After those preprocessing steps, the cone angle is calculated with the proposed methods. Metrics that were used to compare segmented masks with labels were Intersection Over Union (IoU) and Dice Coefficient (Dice). Also, the average time of proposed and published methods was measured. Proposed methods were shown to be faster. There are no ground truth values for cone angles, therefore methods can only be compared mutually. Mean and standard deviations were calculated for each method. While the standard deviation can be considered an indicator of performance, it is not a strong indicator as the cone angle oscillates during spray injection time. A diesel spray image was segmented with three methods. From the segmented image the cone angle was obtained with two proposed and two literature methods. Segmentation methods have no significant difference in metrics. The proposed methods work faster than the literature methods, but accuracy can not be determined due to the lack of ground truth angles.

References:

1 S.N. Soid and Z.A. Zainal. "Spray and combustioncharacterization for internal combustion engines usingoptical measuring techniques – A review", in: Energy36.2 (Feb. 2011), pp. 724–741.

2 L. Bravo and C. Kweon. "A Review on Liq-uid Spray Models for Diesel Engine ComputationalAnalysis (Droplet Breakout)", in:Army ResearchLaboratoryMay (2014), p. 54.

3 I. Ruiz-Rodriguez et al. "Investigation of SprayAngle Measurement Techniques", in: IEEE Access7(2019), pp. 22276–22289.

Heterogeneous Computing and Advanced Cloud Services

On the Inverse Kinematics Analysis with Machine Learning Algorithms

S. Šegota, Baressi, N. Anđelić, I. Lorencin, D. Štifanić, J. Musulin, M. Glučina, Z. Car

University of Rijeka, Faculty of Engineering

The calculation of inverse kinematics (IK) is a key part of industrial robot modeling, as it allows for the determination of the joint angles necessary to place the end-effector in the desired position. Still, analytically determining the IK equations is time-intensive, as it requires the solution of a complex equation system – and only grows more problematic with the increase in the number of joints of the robotic manipulator. As the larger robots require a more complex modeling process for determining the IK equations, solutions are required for easier determination of the IK equations [1]. Machine Learning (ML) is a possible solution to complex issues as described one. Development of the model that can determine IK would allow for faster modeling of robotic manipulators, without the need for error-prone mathematical analysis.

To develop the model using ML, the dataset is created through direct kinematics (DK). Determining DK equations is a comparatively simple process that tells us the positions of the end-effector according to the given set of joint angles. This allows us to generate the joint angle positions uniformly randomly, with each value being selected from the range of motion of individual joints as provided by the robotic manipulator manufacturer. Then, the randomly obtained values are used as variables in the DK equations obtained by the Denavit-Hartenberg process, in order to generate the tool positions. This means that the joint-space coordinates $[q_1, q_2, q_3, q_4, q_5, q_6]$ are transformed into tool-space coordinates including spatial $[x, y, z]$ and orientational $[\psi, \theta, \phi]$ coordinates. This process allows for the creation of the dataset with an arbitrary size. In the presented research, the size of the dataset is 15000 points [1].

To determine the model for the transformation of tool-space coordinates into the joint-space coordinates, in other words, the IK model, the vectors consisting of tool-space coordinates $[x, y, z, \psi, \theta, \phi]$ are used as the input, while each individual joint angle q_i is used as the output. Each output is regressed separately as the methods used can only regress a single value with a single model.

The model of the industrial robotic manipulator used for the presented research is IRB 120, manufactured by ABB [2]. ABB IRB 120 is a popular robotic ma-

nipulator, commonly used in industry for pick-and-place operations when smaller parts are acted upon, such as electronic circuits or smaller machine parts.

The methods used in the research are multilayer perceptron (MLP), which is a feed-forward neural network, and symbolic regression (SR), also known as genetic programming, which is a method that allows for the development of equation-shaped models through the process of computational evolution. MLP consists of an input layer, size of which is six neurons - equal to the number of inputs, one or more hidden layers, and an output layer consisting of a single neuron, which will be the predicted output $\hat{q_i}$, corresponding to the real output q_i [2]. SR creates models by generating a set population of candidate solutions, which are tree-shaped equations that attempt to model the output based on the input values. On this initial set, the evolutionary computing operations are performed. These operations are crossover - which chooses two possible solutions and combines them into one solution, mutation which randomly modifies a candidate solution, and reproduction which copies the candidate solution in the next generation. The candidate solutions on which these operations will be performed are selected through a process called fitness-proportionate selection. This means that the higher quality solutions, which have a better fitness value (lower error), are more likely to be selected. The repetition of this process will enable the increase in fitness of candidate solutions, leading to the convergence towards the optimal solution [3].

Both MLP and SR methods have hyperparameters that determine their performance. For MLP the adjusted hyperparameters are the number of hidden layers, number of neurons in each hidden layer, learning rate type, initial learning rate value, activation function, and L2 regularization parameter. For SR the hyperparameters are the size of the population, the number of generations (iterations in which the evolutionary operations are applied), and the probabilities of each operation occurring. These hyperparameters are selected uniformly randomly across the possible ranges, with the only exception being evolutionary operations probabilities. These are calculated by first selecting the probability of the crossover operation, P_C happening, in the range of $[0.80, 0.95]$. Then, the probability of different mutation types is calculated as $P_H \in [0, 1.0 - P_C]$ for hoist mutation, $P_P \in [0, 1.0 - (P_C + P_H)]$ for point mutation, and $P_S T \in [0, 1.0 - (P_C + P_H + P_P)]$ for subtree mutation. The probability range of reproduction is calculated similarly with $P_R \in [0, 1.0 - (P_C + P_H + P_P + P_S T)]$. This is done because the sum of probabilities of all evolutionary computations happening, including their variants, has to be satisfied per $P_C + P_H + P_P + P_S T + P_R = 1.0$ [3].

The dataset is split into 80:20 training-testing split. As the entire dataset consists of 15000 points. This means that the models are trained using 12000 data points and tested using 3000 data points. Due to a large number of data points

and considering that their distributions are uniform (for outputs) or normal (for inputs) there is no need to apply cross-validation, and a straight training-testing split may be performed. The performance of the models is evaluated using Mean Average Percentage Error (MAPE). If q_i are the real values contained in the testing set, $\hat{q_i}$ are the predicted values obtained from the model, and n is the number of points in the testing set, then $MAPE$ can be calculated as $MAPE = \frac{1}{n}\sum_{i=1}^{n} \mid \frac{\hat{q_i}-q_i}{q_i} \mid$. $MAPE$ was selected because it is a normalized error, allowing us to easily compare performance on the individual outputs when the outputs do not fall within the same range.

After the training and evaluation of models for individual outputs and methods the obtained results show that a quality regression is achieved. For MLP the best achieved results were: $MAPE_{q_1}^{MLP} = 0.018$, $MAPE_{q_2}^{MLP} = 0.014$, $MAPE_{q_3}^{MLP} = 0.025$, $MAPE_{q_4}^{MLP} = 0.037$, $MAPE_{q_5}^{MLP} = 0.008$, and $MAPE_{q_6}^{MLP} = 0.34$. Results achieved using the SR algorithm were: $MAPE_{q_1}^{SR} = 0.092$, $MAPE_{q_2}^{SR} = 0.102$, $MAPE_{q_3}^{SR} = 0.226$, $MAPE_{q_4}^{SR} = 0.152$, $MAPE_{q_5}^{SR} = 0.054$, and $MAPE_{q_6}^{SR} = 0.912$.

The results show that quality regression models have been achieved. The last joint shows the poorest regression, which may be caused by the multiple of similar x, y, and z positions resulting in different placement for the joint q_6 due to this joint only affecting the orientation of the end-effector in the tool-space coordinates. Compared to previous research in the field, it is shown that the inclusion of tool-space orientation coordinates has a slight positive effect on the developed models. The results for SR are poorer than those obtained with the MLP, but still within a satisfying range of error - with even the largest error being less than 1% of the total range, which is within the error ranges provided by the manufacturer. This may be caused by the higher complexity of models obtained from MLP, which has the result of the higher precision regression models. Still, it should be noted that SR generates models which can easily be exported in the shape of the equations, using standard arithmetic and trigonometric operations. This allows for easier integration of SR models into different software, as those models are programming language and library agnostic, unlike MLP generated models. For this reason, a conclusion can be drawn that MLP produces models which are better for tight integration in tasks where high accuracy is necessary, while SR produces models which are more easily applied in a variety of environments.

References:

1 S. Baressi Šegota, N. Anđelić, V. Mrzljak, I. Lorencin, I. Kuric, and Z. Car. "Utilization of multilayer perceptron for determining the inverse kinematics of an industrial robotic manipulator." International Journal of Advanced Robotic Systems 18, no. 4 (2021): 1729881420925283.

2 ABB Group. "Product Specification IRB-120", 1st ed., Zurich, Switzerland
3 Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, J. Musulin, D. Štifanić, and V. Mrzljak. "Determining inverse kinematics of a serial robotic manipulator through the use of generic programming algorithm." Proceedings of 8th International Conference of Serbian Society of Mechanics, (2021), Kragujevac, Serbia, June 28-30, 2021, pp. 1-6

Implementation of Distributed Ledger Technology in Dew Computing applications

V. Bojović

Ruđer Bošković Institute, Centre for informatics and computing, Zagreb, Croatia

Digitalisation is present in today's whole human civilisation. It was crucial to develop network computing which is important for blockchain technology. In the last few years development of Blockchain technology became one of the major new developments. A Blockchain can solve a lot of problems e.g. traceable storage , payment and direct coordination. Dew systems are envisaged to be deployed in wide range of application domains no matter of geographic locations and network availability. This system is designed to work in semi-internet connected infrastructure. Instead of relying on the centralized computing models, dew computing can benefit from peer-to-peer data sharing among the dew droplets. In case when there is no connection between dew and fog or dew and cloud, dew droplet can process it's own part of work and then dew droplets can exchange information between each other and put it on cloud/Internet soon as connection is established. Communication between dew droplet and Internet can be direct or it can be provided by fog/edge who provides preparation, coordination and computation functionalities. Dew droplets store small amount of data. Their storage can be filesystem or database where chain of blocks or decisions can be stored. After dew droplets are done with processing of block of data, that block can be accepted by blockchain when the Internet becomes available. In the Dew-Edge-Fog-Cloud hierarchy, blockchain technology is a promising new approach. This architecture helps overcome latency and resource management problems. That approach is going towards enabling novel applications in a variety of fields, from social and educational to scientific and industrial. In this work HashNet is chosen because it's algorithm is scalable, fast, secure and decentralized. Reason why proof of work is not used in this scenario is because it consumes too much power. Alternative is proof-of-stake which is more secure and consumes less power than PoW. Proof-of-authority (PoA) helps to avoid the possibility that certain stakeholders take over transactions which is possible in PoS, and it uses identity as a stake. Author states that he was participating in

idea creation, and was studying possibilities of connecting dew computing and blockchain.

Computer-aided oral squamous cell carcinoma detection based on AI

J. Musulin, D. Štifanić, A. Zulijani, S. Šegota, Baressi, Z. Car

University of Rijeka, Faculty of Engineering, and Clinical Hospital Center in Rijeka

Oral squamous cell carcinoma (OSCC) is the most common histological neoplasm of head and neck cancers, and while it is located in an easily visible area and can be detected early, this does not usually occur. OSCC is typically treated primarily through surgical resection with or without adjuvant radiation, which has a significant impact on patient quality of life [1]. Despite significant progress in understanding the complex process of carcinogenesis, no reliable tool for prognostic prediction has been discovered. Clinical examination, conventional oral examination, and histopathological evaluation following biopsy are currently golden standard methods for detecting oral cancer, which can detect cancer in the stage of established lesions with significant malignant changes. However, the main issue with using histopathological examination for tumor differentiation, as well as a prognostic factor, is the subjective component of the examination. For that reason, in our study, we are using Artificial Intelligence (AI) algorithms in order to improve objectivity and reproducibility, respectively to reduce inter-and intra-observer variability. Moreover, such an approach could have a direct impact on patient-specific treatment intervention by identifying patients' outcomes. It could also help the pathologist reduce the number of manual inspections as well as make faster decisions with higher precision. The dataset used in the study consisted of 322 immunohistochemistry (IHC) stained histopathological images with a resolution of 768 x 768 pixels. The formalin-fixed, paraffin-embedded oral mucosa tissue blocks of histopathologically reported cases are obtained from the archives of the Clinical Department of Pathology and Cytology at the Clinical Hospital Center in Rijeka. Sample slides were reviewed and classified by two unbiased pathologists, furthermore, the segmentation masks were prepared and validated by another independent clinician. Images are split into three classes: grade I (well-differentiated), grade II (moderately differentiated), and grade III (poorly differentiated) while the ground truth masks consisted of epithelial and stromal tissue. The segmentation of epithelial and stromal regions, as well as the classification of tumor grades, are the two most prevalent tasks in histopathological image analysis. In our study [1, 2] different Deep Learning Models are applied to the same problem in order to investigate the influence on the overall outcome in terms of multiclass classification performance. Experimental results are achieved

with Xception, ResNet50, ResNet101, MobileNetv2, InceptionV3, InceptionResNetV2, DenseNet201, NASNet and EfficientNetB3 architectures which are pretrained on ImageNet. Each model architecture is trained with three optimizers: stochastic gradient descent (SGD), Adam, and RMSprop. Overall results of the [1] are superior, where integration of Xception and SWT resulted in the highest classification values of 0.963 AUCmacro and 0.966 AUCmicro with the lowest standard deviation of $\pm\sigma = 0.042$ and $\pm\sigma = 0.027$, respectively. However, these results are achieved when preprocessing method based on stationary wavelet transform (SWT) and proposed mapping function were used. Moreover, the correct comparison is when performances of the methods without preprocessing are directly compared. In this case, the DenseNet201 architecture in combination with Adam optimizer achieves highest AUCmacro and AUCmicro values of 0.953 ($\pm\sigma = 0.043$) and 0.949 ($\pm\sigma = 0.041$), respectively. These results are obtained when two additional layers are added to the base DenseNet201 architecture. The first layer added was global average pooling, and the second was fully connected which was the output layer [2]. The first step in studying the tumor microenvironment and its impact on disease progression is to segment the tumor into epithelial and stromal areas. Semantic segmentation, also known as image segmentation, is used to group areas of an image that correspond to the same object class. In our study, $Xception_{65}$ was used as $DeepLabv3+$ backbone in order to perform semantic segmentation. $DeepLabv3+$ is a deep learning model that aims to assign semantic labels to any pixel in the input image. The model was pre-trained on the Cityscapes dataset before the training on the OSCC dataset was performed. As model input, original images and SWT approximations were used along with corresponding ground truth masks. From obtained results it can be concluded that integration of $Xception_{65}$ and $DeepLabv3+$ resulted in the highest semantic segmentation values of 0.879 ± 0.027 mIOU, 0.955 ± 0.014 F1 score and 0.941 ± 0.015 accuracy. Classification models along with segmentation model are performed on the immunohistochemistry (IHC) stained histopathological images, however, there is another type of tissue staining that is of particular interest to pathologists called haematoxylin and eosin (H&E) staining. This is because the H stain highlights nuclei in blue against a pink cytoplasmic background (and other issue regions). This allows a pathologist to quickly identify and examine tissue, which is a labour-intensive operation. An automated (H&E) stain histopathological analysis could assist the pathologist in discovering new informative features and in analyzing the tumor microenvironment. In order to perform automated image analysis, H&E-stained images need to be normalized. This is due to the large color variations in images caused by sample preparation and imaging settings. In our study, we used the SVD-geodesic method for obtaining stain vectors where the first step is to convert the RGB color vector to their corresponding optical density (OD) values then remove data with OD

intensity less than β [3]. Threshold value of $\beta = 0.15$ was found to provide the most robust results while removing as little data as possible. The next step is to calculate singular value decomposition (SVD) on the OD tuples then create a plane from the SVD directions corresponding to the two largest singular values. After projecting data onto the plane and normalizing to the unit length we calculate the angle of each point wrt the first SVD direction. The final step is to convert extreme values back to OD space. The findings indicate that automated segmenting epithelial and stromal tissue has a lot of potential for quantifying qualitative clinic-pathological aspects and predicting tumor aggressiveness and metastasis. In future work, segmented sections will be used to study the tumor microenvironment, where the stroma is crucial for epithelial tissue maintenance. AI computer-aided tools for analyzing the tumor microenvironment could help in treatment planning, improving prognosis and survival rates, and ensuring that patients have a high-quality life. AI analyses were performed using Python on a GPU-based High-Performance Computing (HPC) server. The server consists of two Intel Xeon Gold CPUs (24 C/48 T, at 2.4 GHz), 768 GB of ECC DDR4 RAM, and five Nvidia Quadro RTX 6000 GPUs, with 24 GB of RAM, 4608 CUDA and 576 Tensor cores.

References:

1 Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers, 13(8), 1784.

2 Musulin, J., Štifanić, D., Zulijani, A., Baressi Šegota, S., Lorencin, I., Anđelić, N., & Car, Z. (2021). Automated Grading of Oral Squamous Cell Carcinoma into Multiple Classes Using Deep Learning Methods. The 21st IEEE International Conference on BioInformatics and BioEngineering.

3 Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., ... & Thomas, N. E. (2009, June). A method for normalizing histology slides for quantitative analysis. In 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 1107-1110). IEEE.

Regression Models for Geospatial Big Data

D. Katušić, K. Pripužić

University of Zagreb, Croatia, Faculty of Electrical Engineering and Computing, and Centre of Research Excellence for Data Science and Advanced Cooperative Systems

There are different approaches that seek to model heterogeneity present in the spatial data. Spatial heterogeneity expresses different influence of independent

variables on the dependent variable at the observations' locations. These relationships can change abruptly between smaller areas within the total observed spatial area, while within these smaller areas they remain relatively homogeneous. They are important for data visualization and prediction modeling.

Linear regression (generalized linear regression, GLR) spatial models are most widely used to describe spatial heterogeneity relationships. However, the effects of independent variables in such models are often characterized by constant regression coefficients over total observed spatial area which cannot qualitatively capture more complex spatial patterns. Geographically weighted regression (GWR) model was developed in response: it is a global regression model composed of a series of local overlapping models best suited to observations. Therefore, GWR extends the ordinary global GLR model by fitting a specific local regression model to each observation. Such an approach is extremely suitable for coarse data with inherent spatial heterogeneity. When estimating regression model parameters for one location, the influence of observations from that location and nearby observations is combined, typically adding more weight to those closer and less to those more distant. The idea of spatial weighting is a natural strategy used in the light of Tobler's first law of geography („Everything is related to everything else, but near things are more related than distant things."). Weighting strategies are implemented by combining different fixed or adaptive kernel functions (e.g. uniform, gaussian, exponential, etc.) and bandwidth boundaries. Although GWR models achieve better prediction results than GLR models, training them and then performing prediction can result in poor computational performance. The challenge of efficiently extracting knowledge from big data sets is a general problem of all machine learning models. Training a local model for each observation location is a very complex and computationally expensive operation in the case of a big data set. For GWR model this can lead to an excessive number of model parameters. Over time modifications of the basic GWR model were developed. Some of them, usually denoted as geographically time weighted regression (GTWR) models, add time dimension and for some applications provide even better prediction results. However, they do not address the problem of computational complexity for big data sets. For application of GWR or GTWR in real-time, the running time complexity of these models needs to be reduced to approximately O(1). Possible way to achieve this is to develop prediction models for spatial regions in advance and train a fixed number of local models, thus avoiding training local models anew each time in the prediction phase. One of the ways to adequately determine spatial regions in which observation locations are grouped is the application of clustering algorithms.

To test different approaches of efficient modelling of spatial heterogeneity, the following models have been implemented and evaluated: classical multiple lin-

ear regression that models dependent variable as a single linear function of several independent variables (GLR), geographically weighted regression (GWR) model proposed in [1], geographically time weighted regression (GTWR) model described in [2], its time-only based variant (TWR), and our custom implementations of geographically clustered regression (GCR) and geographically time clustered regression (GTCR) that adds time dimension to enforce better predictions. Both latter models were inspired by the conclusions from paper [3]. All models are implemented in two variants: centralized that run on a single computer and distributed that run on multiple computers in a cluster. The first are based on Python, the latter are based on Apache Spark framework using the Spark MLlib (https://spark.apache.org/docs/latest/ml-guide.html) and Sedona (https://sedona.apache.org/) libraries. Data set used for model evaluations is Beijing Multi-Site Air-Quality Data Set (https://archive.ics.uci.edu/ml/datasets/Beijing Site+Air-Quality+Data [420k observations and 14 features (PM2.5, PM10, SO2, NO2, CO, O3, TEMP, PRESS, DEWPNT, RAIN, WINDSPD, lat, long, timestamp)]).

Developed models use PM2.5 as dependent and all descriptive features as independent variables. The reason is the following: if there is a good enough model that can predict the concentration of PM2.5 particles based on the concentrations of other measured particles, there is a possibility of significant savings on air quality measurement equipment. Models have been evaluated for their accuracy in prediction modelling (first outlined problem) and computational performance on big data sets (second outlined problem). First metric used for evaluating prediction is the root mean square error (RMSE). It shows differences between the observation values (y_i) and the values predicted by the model ($\hat{y}_i$). Smaller value of RMSE means that the data are closer to the prediction line, thus the prediction error is smaller. Second prediction modelling metric is coefficient of determination (R^2). It gives the percentage of variation in the observations (y_i) explained by the model ($\hat{y}_i$). The range goes from 0 to 1 and the goal is to achieve the highest possible value. Both metrics display identical results for all implemented models: GLR enables poorest predictions, while GTWR and GTCR achieve the best. Measurement of the overall execution time (training and prediction phase) of each of the models was performed on a set of approximately 40,000 observations. There is a difference in the training phase time between weighted regression (GWR, GTWR) and clustered regression (GCR, GTCR) models, as the former have shorter training phase. However, the results for the prediction phase show the opposite: weighting-based models require much more time to predict than clustering-based models. As explained in the chapter "Problem description", weighting-based models need to retrain models anew in the prediction phase, which results in the explosion of computational complexity. Consequently, this inspired usage of clustering as an optimization technique to bypass this issue.

The execution times of the GWR and GTWR with fixed and variable kernels are approximately equal and increase sharply with increasing size of the evaluation set. On the other hand, the execution times of different clustering-based models increases linearly with increasing size of the evaluation set and mostly depend on the number of clusters determined as a hyperparameter of the model. To summarize, weighting-based models (GWR and GTWR) cannot be used to process big data sets in approximately real-time, while clustering-based models (GCR and GTCR) are more applicable in this case.

Centralized and decentralized versions of all mentioned models have been implemented and their performance evaluated: best prediction results are achieved by GTWR and GTCR. Best computational performance results are achieved by: GTWR (training phase), GCR/GTCR (prediction phase and overall). Possible future work directions include implementing additional models for prediction modelling comparison and trying new optimization methods for model computational performance not necessarily based only on clustering methods.

This work has been supported in part by Croatian Science Foundation under the project UIP-2017-05-9066. This research has been supported in part by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

References:

1 A.S. Fotheringham, C. Brunsdon and M. Charlton: "Geographically Weighted Regression: The Analysis of Spatially Varying Relationships", Wiley, 2002

2 A.S. Fotheringham, R. Crespo and J. Yao: "Geographical and Temporal Weighted Regression (GTWR)", Geographical Analysis, vol. 47, no. 4, pp. 431-452, 2015

3 S. Sugasawa and D. Murakami: "Spatially Clustered Regression", Spatial Statistics, vol. 44, 2021

Detecting and filtering unwanted features from GPR recordings using stationary wavelet transform

D. Štifanić, J. Musulin, S. Šegota, Baressi, M. Glučina, Z. Car

University of Rijeka, Croatia, Faculty of Engineering

Detecting and identifying underground objects without excavation can be demanding, time-consuming, and at the same time, very challenging. Nowadays, with different approach i.e., by utilizing ground-penetrating radar (GPR), underground utilities such as pipes, metals, cables, concrete etc. can be observed and explored from the ground surface [1]. Since such a method is nondestructive,

sub-surface surveying has been achieved using electromagnetic radiation. First, a high-frequency electromagnetic wave is emitted into the ground where the signal is reflected, scattered, or refracted from different subsurface structures. Afterwards, the returning signal is received and recorded by the GPR. The received signal can be visually interpreted as an image where the x-axis represents the position of the radar, while the y-axis represents the depth of wave penetration. After the ground has been observed, and recordings acquired, the next step is analyzing and post-processing the data. In real-world applications, GPR recordings can be very noisy due to the presence of different unwanted objects or specific soil layers that can produce interference. Such unwanted features are contained in the same recordings as the valuable features of the observed objects. However, using a mathematical tool called Wavelet Transform (WT), the unwanted features can be partially filtered out [2]. By applying WT on the image obtained using GPR, the data can be decomposed into multiple frequency bands, thus simplifying the analysis and interpretation process. The decomposition process itself results in approximation and detail coefficients. However, when dealing with two-dimensional data, the detail coefficients consist of horizontal (HL), vertical (LH), and diagonal (HH) features of an image. The decomposition process can be repeated until specific requirements for analysis are met. In order to filter out problem-specific features, a variety of wavelet functions can be utilized, where each one of them excels in different properties. The main aim of this research is post-processing GPR data in terms of filtering out features of an image that are captured as reflections of different soil layers or other objects which have differences in physical properties of materials. Such features can be contained in all three types of detail coefficients. Since in this research the object of interest that is being observed is underground infrastructure, more precisely, metal and PVC pipes, valuable information are hyperbola-shaped features contained within the GPR images. Filtering out unwanted features is the first step of processing i.e., preparing the data for detection and localization of underground infrastructure using machine learning-based models. The conventional discrete wavelet transform (DWT) suffers in terms of decimation and shift-invariance, thereby, for the purpose of this research and to overcome the aforementioned drawbacks, stationary wavelet transform can be utilized. This way, shift-invariance, no decimation performed after each decomposition level, and better overall time-frequency localization are ensured, which makes this approach feasible for pattern recognition, and change detection [3]. In order to analyze each subband separately, the experimental image captured with GPR was decomposed utilizing the SWT with Reverse biorthogonal (rbio) 1.1 wavelet function at five decomposition levels. This way, the approximation and detail coefficients (HL, LH, and HH) were obtained for each decomposition level separately resulting in twenty images of features. Afterwards, by analyzing the detail coefficients it can be seen that, in

order to filter out unwanted features, HL, LH, and HH coefficients thresholds must be determined according to the level-dependent estimation of the intensity of unwanted features. Considering the aforementioned, a satisfactory result in terms of filtering out the features within GPR recordings were achieved when hard thresholding was utilized for all detail coefficients at all five decomposition levels. Furthermore, the values of thresholds at each level were as follow: level 5 (HL – 1008, LH – 1008, HH - 1008), level 4 (HL – 453.2, LH – 453.2, HH – 453.2), level 3 (HL – 132.4, LH – 132.4, HH – 132.4), level 2 (HL – 101.6, LH – 101.6, HH – 101.6), level 1 (HL – 43.24, LH – 43.24, HH – 39.05). According to the obtained results, the findings indicate that post-processing of GPR recordings using SWT can significantly contribute to future work where artificial intelligence-based algorithms will be utilized for detection and localization of underground infrastructure based on the GPR recordings.

References:

1 Jol, H. M. (Ed.). (2008). Ground penetrating radar theory and applications. Elsevier.

2 Guo, X., Li, Y., Suo, T., & Liang, J. (2017). De-noising of digital image correlation based on stationary wavelet transform. Optics and Lasers in Engineering, 90, 161-172.

3 Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers, 13(8), 1784.

Experimental study of the impact of feature selection and resampling methods on software defect prediction

M. Dudjak, G. Martinović

Faculty of Electrical Engineering, Computer Science and Information Technology, Osijek

Software Defect Prediction (SDP) is one of the most useful activities during the testing phase of the software development life cycle which identifies modules that are prone to failure and require additional testing. Though SDP is a helpful testing method, various issues make it difficult to predict the defective modules, such as class imbalance, class overlapping, and small disjuncts. This experimental study analyses the contribution of feature selection (FS) and resampling techniques used to mitigate the adverse effects of these characteristics. Recommendations for approaching the SDP problem are made based on the obtained results. All datasets employed in the experimental study originate from the PROMISE repository [1] and represent publicly available NASA datasets

for predicting software defects [2]. All datasets represent binary classification problems, while features are derived from software complexity metrics such as lines of code measures, McCabe's metrics, Halstead's metrics, and others [3]. A total of 12 datasets were selected to comprise the test bed. In order to substantiate the presence of the mentioned difficulties, the absolute values of the data complexity metrics were derived. Although it is ill-advised to consider the mere absolute values of data complexity measures, the problems of class overlap and small disjuncts manifest within SDP datasets, as in almost no aspect is a minimum degree of complexity obtained. However, it is difficult to determine the actual degrees of these intrinsic characteristics, as the values of these metrics are intended to be compared with the same measures on other problems or used to derive correlation with classification performance as done later in this paper. To assess the behaviour of several well-known classification algorithms when faced with SDP task, a comprehensive experimental analysis was conducted. The classifiers that have been analysed and compared are decision tree (DT), k-nearest neighbours (k-NN), Gaussian naive Bayes (GNB), multilayer perceptron (MLP) and support vector machine (SVM). It is important to note that the parameter tuning was performed for all algorithms, to provide a fair comparison. The obtained results coincide with the findings from the literature, according to which Naive Bayes is a suitable algorithm for the SDP task. It is undoubtedly the best-performing classifier in terms of minority class prediction performance, while it lags far behind in recognising the majority class. However, the absolute values of the AUC results achieved are very low, suggesting that all classifiers have problems with the correct classification of a significant proportion of instances from employed datasets. The results presented suggest that class imbalance in combination with other characteristics of SDP datasets makes learning difficult. In general, class imbalance can be declared the most detrimental characteristic of SDP datasets, followed by class overlap, while the negative impact of small disjunct is noticeable only for the GNB classifier. It is logical to assume that reducing the degree of these characteristics would reduce their negative impact and thus increase the classification performance. According to the literature, the removal of overlapping feature values can be achieved by performing feature selection, while class imbalance and small disjuncts can be alleviated with either oversampling or undersampling procedures. After conducting the FS procedure, the changes achieved in performance are not significant (on some datasets, performance even deteriorates), but what is very interesting is the exceedingly high reduction rate. It suggests that most software metrics within employed datasets do not contribute to the construction of classification models. Given that there are no excessive differences in performance, feature selection can be declared a helpful step, at least for significant dimensionality reduction of SDP datasets. However, taking into account the fact that FS eliminates $65-85\%$

features from SDP datasets, while inconsiderably affecting classification performance and feature overlap measures, one may conclude that a significant degree of class overlap remains in the remaining $15 - 35\%$ features. Based on the literature review, resampling algorithms are the most common choice to mitigate the adverse effects of the described data intrinsic characteristics. For this purpose, two well-established resampling algorithms were selected, namely synthetic minority oversampling technique (SMOTE) and edited nearest neighbours (ENN). The neighbourhood size k in SMOTE and ENN resampling procedures was set to 5, and algorithms were executed until the number of instances of both classes was equalised ($IR = 1$). Both algorithms increase the performance of all utilised classifiers, with SMOTE significantly outperforming ENN. Additionally, after performing resampling procedures, there are no statistically significant differences between the observed classifiers. Given that these procedures eliminate class imbalance from SDP datasets, and that redundant and detrimental features are removed via FS, the results suggest that the problem of class overlap remains, making it difficult to further improve performance. The conclusion that can be drawn from the presented results is that class imbalance and other data intrinsic characteristics cause algorithms to behave like a majority class classifier. Moreover, no software complexity metric used as a feature in the classification model manages to unambiguously separate minority and majority instances. Such a distribution of feature values explicates the exceptionally high reduction rate after employing the feature selection technique. Nevertheless, FS fails to significantly improve the performance of the observed classifiers but can still be considered a useful step as it results in a significant dimensionality reduction. On the other hand, performing resampling can significantly reduce the adverse impact of class imbalance and small disjuncts and generally improve the performance of all considered algorithms. In doing so, oversampling proved to be more effective than undersampling. Therefore, the application of the oversampling procedure is to be recommended when dealing with the SDP task. Finally, in future work, efforts will be made to evaluate or design additional metrics for measuring levels of studied data intrinsic characteristics in SDP datasets to advance the translation of the results obtained in this study. Until then, we hope that these results can serve as a kind of general framework for selecting the appropriate algorithm for the SDP problem.

References:

1 J. S. Shirabad, T. J. Menzies et al., "The promise repository of software engineering databases," School of Information Technology and Engineering, University of Ottawa, Canada, vol. 24, 2005.

2 M. Chapman et al., "Metrics data program, nasa iv and v facility", http: mdp.ivv.nasa.gov., 2004.

3 Y. Ma et al., "Data complexity analysis for software defect detection," In-

ternational Journal of Performability Engineering, vol. 14, no. 8, p. 1695, 2018.

VINI 4.0: Virtual drug screening tool based on biological networks of cancer

J. Mesarić, D. Tomić, D. Davidović

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, and Ruđer Bošković Institute, University of Zagreb, Croatia

Although *in-silico* models of cancer are already valuable tools for detection of new drugs and for the establishment of more precise cancer diagnoses, they are still not accurate enough to be implemented in clinical practice. Besides their limited accuracy, they impose a high demand on computational resources. Increasing the *in-silico* model accuracy and speed are the most important goals, and this can be accomplished by using larger and more complete datasets, by personalized approach to each cancer patient, and by using well developed mathematical algorithms which solve large optimization problems. Optimal solution means more precise diagnosis and more effective therapy. Virtual drug screening is one of the most widely used approaches for finding new potential drugs. The process consists of selecting one or more chemical compounds with the highest binding free energy to target proteins. Up to recently, the vast majority of proteins still lacked the experimentally obtained 3D structures required by most structure-based analysis software. This has been solved with AlphaFold [1] which is a neural network trained to predict 3D protein structure from its sequence with atomic precision. Before DeepMind we had to use online services like SWISS which is a database containing already predicted structures and with on-demand protein homology-modelling to obtain its 3D structure from a given sequence. This allows us to easily get structures of new and unseen proteins and use them for binding energy calculation.

Based on additional information on chemical compounds whose anti-cancer effectiveness needs to be estimated, Vini transforms cancer pathways into binding energy matrices and subsequently calculates second largest eigenvalue (SLEM) values of these matrices. Binding energy matrices describe the interaction of cancer with chemical compounds. Vini [2] model now integrates external databases containing gene expression from clinical trials, enriching the model with data obtained by mass spectrometry and RNA-seq methods. Gene expression profiles were added to the Vini model in order to get higher accuracy of predictions by augmenting structure-based data with proteomic analysis. Gene expressions for each biological network can be used to better describe the importance of each specific gene in that biological process.They can be introduced in the energy

binding matrix as gene expression factor. We are multiplying EB matrix diagonal elements with value of 1 if there is no expression for that particular gene. Including proteomic data is a step towards personalized medicine because now we can use the proteomic profile of specific patients to get more accurate predictions of binding energies.

Unlike most virtual drug screening tools, Vini has the ability to screen double and triple combinations of drugs against a given target. We ran binding energy calculations for breast cancer KEGG pathway (map05224) with included gene expressions from MDA-MB-231 MDA-MB-436 cell lines for single and double combination of different FDA approved drugs as ligands. Comparison of results obtained with *in-silico* method and *in-vivo* method is currently in process. Binding energies of different approved drugs were calculated so now we can use them to predict growth rate (IC50) of those specific cell lines - MDA-MB-231, MDA-MB-436. Those breast cancer cell lines are available for testing at the Laboratory for Hereditary Cancer (RBI) and further research follows to determine if there is a correlation between the number of entities in KEGG cancer pathways and the obtained Pearson correlation coefficient values. We expect to confirm our hypothesis that there is a significant positive correlation between them. Increasingly popular treatment of cancer is monoclonal antibodies (MAB) which is essentially a drug in a form of protein, which is always a very large molecule compared to "traditional" small molecule ligands (>900 Daltons). Calculation of binding energies for protein-protein pairs is more complex and computationally demanding compared to protein-ligand interactions due to different dockings and scoring mechanisms. We have identified those screening pairs as computational bottlenecks of Vini and proposed a prediction model based on a convolution neural network and a selection of specific features to overcome the existing limits of the current scoring mechanism. Past research has shown that it is important to use datasets that are large and diverse enough to cope with the problem of ligands and protein structure similarity, as well as to avoid overfitting by extracting only representative protein properties. As many proteins share very similar structures, it is easy to overtrain a model by using a dataset that contains too much similar structures causing the loss of the model's generalization ability. Dataset with protein-protein pairs was created from various FDA approved MABs paired with known proteins connected to a biological network of specific cancer. Since in this case both MAB drugs and their targets are large molecules we can not use standard tools for protein-ligand docking like AutoDock VINA. For this purpose we have used Rosetta Common tools which is an integrated solution for creation of protein-protein complexes and calculation of their binding energies. Binding energies for non-small cell lung cancer are already calculated for the EKVX cell line which is augmented with available data and with new data in future.

After proteins were prepared for docking, sanitized and their binding energies were calculated we can use them as input points in our dataset for training a CNN to predict binding energies for unseen protein complexes. We are using an Atomic Convolution neural network [3] with two hidden layers and an output layer containing a single node. It is a structure-based bioactivity prediction method which uses 3D spatial convolution of atoms in protein structures to learn representation of atomic chemical reactions. This method is initially developed for protein-ligand binding energy prediction from a given crystal structure but instead of a small molecule for a ligand we are feeding it protein. In order to prepare input data for a convolution model we have to create a molecular featurizer describing protein 3D structure as a graph with atoms represented as nodes and interatomic connections as edges. For each molecule we are calculating its vector fingerprint by finding all neighbour nodes up to a defined threshold limit, removing hydrogen atoms, creating matrix with positions of all atoms and creating a z-matrix which contains the atomic number of every atom. Setting a neighbour cutoff limit allows the model to be scaled as some input graphs can contain more than 1000 nodes - this way only neighboring atoms are taken into consideration with an assumption that physically further atoms have much lower impact on each other.

References:

1 Jumper, J et al. Highly accurate protein structure prediction with AlphaFold. Nature (2021).

2 D. Tomić, K. Skala, L. Kranj;ević, B. Pirkić, S. Stifter, I. Smit, Evaluation of the Efficacy of Cancer Drugs by Using the Second Largest Eigenvalue of Metabolic Cancer Pathways, J. Comput Sci Syst Biol, (0974-7231) 11 (2018), 3; 240-248, doi:10.4172/jcsb.1000280

3 Gomes J, Ramsundar B, Feinberg E, et al. Atomic convolutional networks for predicting protein-ligand binding affinity. ArXiv Preprint ArXiv:1703.10603,2017.

Machine Learning and Data Mining Techniques

Autoencoder Network for Multi Illuminant Color Constancy

D. Vršnak, I. Domislović, M. Subašić, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Color constancy is an essential component of the human visual system. It enables us to discern the color of objects invariant to the illumination that is present. However, this ability is difficult to reproduce in software as the underlying problem is ill-posed, *i.e.* for each pixel in the image we only know the RGB values which are a product of the spectral characteristics of the illumination and the reflectance of objects, as well as the sensitivity of the sensor. In order to combat this, additional assumptions about the scene have to be made. These assumptions can either be hand crafted or learned using some deep learning technique. Nonetheless, most of this assumptions only work for single illuminant problems. In this work we propose a novel method for learning these assumptions for multi illuminant scenes using an autoencoder. The network is trained to reconstruct the original image, by first creating the representation of the image where the per-pixel illumination estimation and the canonically illuminated image. The reconstructed images is produced by multiplying these two outputs. These outputs are used to improve the training of the model by encouraging it to create an valid canonical image and thus learn to separate the color of the illumination from the colors of objects in the scene. However, for testing the canonical image and the reconstructed image were discarded and only the illumination estimation was kept. It can be used as a multi illuminant estimation or as an input to some clustering method to create a segmentation of the scene based on illumination. We experimented with both of these outputs and show that our approach increases segmentation accuracy and reduces the number of artifacts in the illumination estimation compared to the same models which were trained to only estimate the per-pixel illumination maps.

All three of the outputs of the network, along with the groundtruth information from the dataset, can be used to compute the composite loss L used in training. For the illumination estimation loss we propose a new regularized loss function designed to preserve the gradual transition between illuminations in realworld scenes. This loss is composed of the mean squared error term between the predicted and actual illumination for each pixel, as well as a regularization term which ensures smoothness. This regularization term is computed by passing the

"uniformity" filter of size $n \times n$,

$$f_n = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ \vdots & \ddots & (n^2-1) & \ddots & \vdots \\ -1 & -1 & \dots & -1 & -1 \end{bmatrix}.$$

For both the reconstruction and the extracted canonical image mean squared error is used without any regularization, because, unlike in the case of the illumination, we want to preserve details about the edges and sharp transitions in color. The final loss function is computed as:

$$L(I_p, C_p, R, I_r, C_r, O) = \alpha MSE(C_p, C_r) + \beta MSE(R, O) + \gamma MSE(I_p, I_r) + \delta \frac{1}{M} \sum_{i,j} (f_n * I_p$$

where I is the per pixel illumination map, C is the canonical image, R is the reconstruction, O is the original image, $\alpha, \beta, \gamma, \delta$ are coefficient used to tune the importance of each component to the overall loss.

The method was trained on realworld multi illuminant images with two illuminants, and on artificially relighted single illuminant images to test the performance on images with more than two illuminants. We conducted experiment and ablative study to test the effectiveness of our proposed model. Additionally, image segmentation based on the per-pixel estimations was done using simple K-means clustering because of its speed and easiness of determining the number of illuminants. Our results show that when training illumination estimation models improves both the estimation and segmentation accuracy, and achieves better results than currently existing models for multi illuminant estimation. We also show that the visual fidelity of the estimated illumination map is better when the regularized loss is used.

The p-curve Method for Ilumination Estimation

V. Stipetić, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

When an image is captured, the color of the objects in it is determined not only by their intrinsic properties, but also by the properties of the light source which illuminates the scene depicted in the image. However color can be a useful factor for tasks such as object detection, object recognition or scene segmentation, both for human visual system and for artificial systems in computer vision. For this reason, the human visual system has the ability to perceive the object colors as they would appear under neutral illumination, sometimes referred to as white

light. This ability is called color constancy. Computational color constancy is a computer vision problem that aims at replicating this human visual system ability, and remove the influence of the color of illumination from an image.

The model for image formation in a camera, given for instance in [1] is

$$I_c(\mathbf{x}) = \int_S J(\lambda, \mathbf{x}) E(\lambda, \mathbf{x}) \varphi_c(\lambda) d\lambda \tag{1}$$

here, $c \in \{R, G, B\}$, I_c is channel c of the captured image, $J(\lambda, x)$ is the function that described how much of light with wavelength λ the object at point x reflects, $E(\lambda, x)$ describes the intensity of the illumination at point $\mathbf{x}$ with wavelength λ and φ_c is a function that describes the sensitivity of the camera sensor for specific wavelengths. Usually there are three types of sensors in a camera, one for each channel. The integral is taken over S, which is the entire spectrum of visible light. As the problem of inverting this equation in its most general form is very ill-posed, a common assumption that is made is that illumination is constant across the entire scene, or in other words $E(\lambda, \mathbf{x}) = E(\lambda)$. Another approximation can be made, based on the assumption that the functions φ of all of the sensors have small overlap in the supports, and are large only close to a specific wavelength. This justifies the approximation $\varphi_c(\lambda) = \delta(\lambda - \lambda_c)$, where δ is the Dirac delta function, and λ_c is the wavelength that sensor is the most sensitive to. Using both of those assumptions, the equation 1 becomes

$$I_c(\mathbf{x}) = J(\lambda_c, \mathbf{x}) E(\lambda_c) \tag{2}$$

So, denoting $E(\lambda_c) = E_c$, the illumination estimation problem becomes finding E_R, E_G and E_B. It can be shown, for instance in [2], that it is sufficient to know the direction of the vector (E_R, E_G, E_B) for computational color constancy. Because of this, for purposes of this paper, we will look for that vector normalised to $E_R + E_G + E_B = 3$, so that white light, or neutral illumination, is represented by $(1, 1, 1)$.

The most commonly used methods for color constancy are based on low level natural image statistics. The first one is known as white patch method, in which the assumption is that the brightest parts of each channel in the image would be pure white under neutral illumination. Mathematically, this can be written as the equation

$$||I_R||_\infty = ||I_G||_\infty = ||I_B||_\infty$$

Another commonly used method is the gray world method based on the assumption that the image under natural illumination is, on average, gray. Mathematically, this is phrased as

$$||I_R||_1 = ||I_G||_1 = ||I_B||_1$$

A third method, which generalizes the two previously mentioned ones, is based on p norms of the channels of the image. In the extremes of choices of p we get the gray world method for $p = 1$ and the white patch method for $p = \infty$. The mathematical formulation of this assumption called shades of gray is

$$||I_R||_p = ||I_G||_p = ||I_B||_p$$

One major downside even of the shades of gray method is that it still only uses a single value of p. To avoid this information loss, we propose the p curve method in which the estimate of the illumination color is based on all values of p within a chosen interval. This is done by generating a curve parametrized by choices of p, defined as

$$f(p) = (||I_R||_p, ||I_G||_p, ||I_B||_p)$$

and the choosing the illumination color which minimizes the average distance of that curve from the line $x = y = z$.

$$_{x+y+z=3} \int_a^b d\left(\left(\frac{f_1(p)}{x}, \frac{f_2(p)}{y}, \frac{f_3(p)}{z}\right), l\right) |g(p)|dp$$

For degenerate curves given by a single point, this method projects that point onto the line $x = y = z$, and so is equivalent to the shades of gray method. For curves consisting of larger number of points, the additional information stabilizes the shades of gray method improving estimates and reducing the variance of estimation.

References:

1 K. Barnard, V. Cardei and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," in IEEE Transactions on Image Processing, vol. 11, no. 9, pp. 972-984, Sept. 2002, doi: 10.1109/TIP.2002.802531.

2 A. Gijsenij, T. Gevers and J. van de Weijer, "Computational Color Constancy: Survey and Experiments," in IEEE Transactions on Image Processing, vol. 20, no. 9, pp. 2475-2489, Sept. 2011, doi: 10.1109/TIP.2011.2118224.

Explainability of classification models in the cybersecurity domain

A. Šarčević, A. Krajna, D. Pintar, M. Vranić, M. Šulc

Faculty of Electrical Engineering and Computing, University of Zagreb

Recently, we have been witnessing more and more attacks on information systems. The goal of these attacks is mainly data theft, reputational damage, disabling work and material gain. Global networking, the increasing complexity of computer infrastructures and the continuous migration of many private and business aspects into the electronic domain make the task of maintaining the security of information systems increasingly demanding. Therefore, it is necessary to use the latest technologies based on data analysis, machine learning and artificial intelligence in the cybersecurity domain. Cybersecurity describes the practice of protecting and restoring networks and information systems from actions that violate the basic security requirements of confidentiality, integrity, availability, and authenticity of data or services.

In the last decade, due to the rapid increase in computer power and the availability and ability to store large amounts of data, there has been a significant boost in machine learning. Due to advances in machine learning and autonomous decision making, systems based on machine learning have achieved superhuman performance. Great development can be seen in the field of image recognition, speech analysis, strategic planning, healthcare, and even in the cybersecurity domain. Modern systems for defending information and network systems are increasingly integrating machine learning methods as a way of detecting attacks.

In parallel with the development of increasingly complex models of machine learning, the problem of their transparency is also emerging. Revisions to machine learning model results are essential to better understand the system, to investigate false positive and false negative alarms, and to find system errors or biases. Systems whose decisions cannot be interpreted are hard to believe, especially in high-risk sectors such as healthcare and the cyber-domain. The explainability and interpretability of machine learning models increases confidence in the decisions made by the system. Much of the scientific literature is still focused on improving model performance, and the semantic gap in such models remains a major problem. In high-risk domains such as the cybersecurity domain, system explainability and interpretability are necessary, and often as important, as model performance. In such domains, the algorithms for explaining and interpreting the model must be implemented as separate modules (separate from the machine learning algorithm) and cannot in any way impair the performance of the implemented model.

References:

1 T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal loss for dense object detection", in Proceedings of the IEEE international conference on computer vision (ICCV), 2017, pp. 2980-2988

2 D.P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", Computer Research Repository (CoRR), abs/1312.6114., May 2014.

3 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Nets", in Advances in Neural Information Processing Systems 27 (NIPS), 2014, pp. 2672-2680

Human detection in aerial images for search and rescue operations

N. M. K. Dousai, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Object detection is one of the most researched areas in computer vision. It is the process of determining where exactly the object is in the scene or image and what object has been detected. Object detection refers to finding different types of objects in the scene such as peoples, cars, animals, or other existing objects present in the scene. While normal ground-to-ground imagery has yielded promising results in object detection, detecting objects in aerial imagery is still considered a difficult task. One such important task is to rescue people in search and rescue (SAR) operations from aerial images without loss of life. SAR operations are conducted in wide-open spaces, such as mountains, lowlands, cities, disaster scenarios and marine rescue. Object detection in aerial images depends on several factors such as altitude, low visibility, the object-of-interest, variations in pose and scale, camouflaged environment with rocks and trees, and high-resolution aerial images. In general, search and rescue operations need to be conducted as quickly as possible to identify missing persons. It can be very expensive and use different types of actions such as sending people, sniffer dogs and different types of ground and air vehicles such as cars and helicopters.

In our research, we propose a deep learning-based approach for human detection in aerial images of Mediterranean mountain landscapes from Balkan regions captured using Unmanned Aerial Vehicle (UAV), used in Search and Rescue (SAR) operations. Detecting humans in aerial images is still a complicated task due to the challenges like camouflaged environment and the varying altitude. In most of the aerial images captured from UAV, the humans cover a negligible part of the image of around 0.1-0.2%. To address this problem of small coverage of interest of object we propose ensemble learning based on EfficientDET [1] with Bi-FPN and FC-FPN for HERIDAL dataset. The training results are

well-plotted and tested on different sizes of the regenerated HERIDAL [2] dataset to show the differences in training computational time and accuracy.

References:

1 Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." arXiv preprint arXiv:1905.11946 (2019)

2 Božić-Štulić, Dunja, Željko Marušić, and Sven Gotovac, "Deep learning approach in aerial imagery for supporting land search and rescue missions." International Journal of Computer Vision 127.9 (2019), pp. 1256

Finding Hamiltonian cycles with graph neural networks

F. Bosnić, M. Šikić

*Faculty of Electrical Engineering and Computing, University of Zagreb, and Genome Institute of Singapore, A*STAR, Singapore*

As deep neural networks proved successful in solving variety of task, it is tempting to test them on solving hard combinatorial problems. Research in this direction has up to now focused mostly on the 2-dimensional traveling salesman problem, but there are many similar problems, such as the Hamiltonian cycle problem or the clique problem, on which such methods would be desirable. These problems are mostly NP-hard and it would be unreasonable to expect neural networks, which are of polynomial complexity, to produce exact solutions.

However, from practical perspective it often suffices to find an approximate solution of a particular subproblem. For example, the process of genome assembly, in which a DNA molecule is to be assembles from numerous small, overlapping fragments, essentially reduces to the Hamiltonian cycle problem. But one is interested in solving this problem only for the family of graphs that can be produced during sequencing of DNAs. Naturally, such graphs have certain structural and statistical properties which could aid in solving of the problem. In addition, finding cycles covering say 90% of the graph would also be considered a success. While there are several heuristic algorithms for solving the general Hamiltonian cycle problem, better heuristic can surely be designed specifically for solving the problem of DNA assembly. Many such heuristics are used in state of the art assembly programs [1], but creating them requires great amount of effort and a profound understanding of the problem. With the help of machine learning on the other hand, these heuristics could be *trained* instead.

The assembly problem is indeed a motivation for our work, but we present a study of the Hamiltonian cycle problem on a special class of random graphs. We give two neural network solvers for this problem based on message passing [2] and graph attention [3] neural networks. Both of these are special kinds of

graph neural networks which we believe are key for handling the Hamiltonian cycle problem. Graphs are processed in an auto-regressive manner and neural networks output the next node in the cycle at every iteration. Note also that the fundamental difficulty in the Hamiltonian cycle problem comes from the topological structure of the graph (i.e. its edge structure), as opposed to the traveling salesman problem in 2 dimensions (cities are represented as coordinates in the plane and one is allowed to travel between each pair of cities) where the difficulty comes from the metric structure of Euclidean plane. One can think of these two as complementary subproblems of the general traveling salesman problem.

To examine generalization properties of our neural network models we devise a specific class of random graphs based on theoretical results regarding Erdős-Rény graphs. When sampling a graph G of size n from this class,

$$\mathbb{P}(G \text{ has at least one Hamiltonian cycle}) \xrightarrow{n} c$$

for some fixed constant $0 < c < 1$. In fact, experimental results show asymptotics to be quite stable for $n > 20$. Therefore, we expect the Hamiltonian cycle problem to be of the *same difficulty* no matter the size of graph G. This gives the model a fair chance of generalizing to larger sizes.

References:

1 R. Vaser and M. Šikic. Time- and memory-efficient genome assembly with raven. Nature Computational Science, 1(5):332–336, May 2021.

2 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In D. Precup and Y. W. Teh, editors, Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pages 1263–1272. PMLR, 06–11 Aug 2017

3 P. Velickovič, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks, 2018.

System for product recognition on shelves images

L. Budimir, Antonio, B. Filipović, F. Šikić, S. Lončarić, M. Subašić, Z. Kalafatić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Detection and recognition of products on store shelves has gained increasing attention in the computer vision community in recent years. Deployment of an efficient detection and recognition system could be beneficial not only to retailers but also to customers and companies which supply the stores with their products. Retailers can use such a system to increase sales by detecting misplaced

products or absence of products (so-called out-of-stock situations) on shelves. Furthermore, this system can be used as a guidance for visually impaired customers or as a tool that product companies can use to check whether the shelf share in a certain store complies with the agreed display arrangement (known as planogram) or not. Our work on this project consists of three different subtasks which will be used to improve existing approaches to this problem: virtual scenes generation, classification improvement, and out-of-stock detection.

For the purposes of product recognition and detection, a large enough dataset needs to be collected. However, this process is arduous and very labor-intensive, because one needs to physically visit many different stores (of various retailers) in order to introduce variance into the dataset. Additionally, there are multiple product categories in each store, all of which need to be represented in the dataset. Therefore, we have decided to generate virtual scenes using the Blender software and render images of them, which corresponds to shooting pictures of real scenes. We have already developed several Python scripts which manipulate the virtual shelf in various ways, such as stocking the shelf, clearing the shelf, simulating the shoppers' behaviors by employing Markov chains, rotating the products randomly, and rendering the images themselves. For future research, we plan to train a model on the generated dataset, which will be used to detect product orientations. Additionally, we plan to introduce more realistic textures into our dataset to improve the performance of the future model.

Product recognition is a large-scale classification problem because one store can display several thousand different products on its shelves [1]. Furthermore, retail stores can add or remove new products daily, and the product's visual appearance is often subject to various changes due to seasonal packaging or rebranding. The product recognition algorithm also must distinguish minor differences in labels and packaging between products from the same category or distributor. These challenges make this task slightly more complex in comparison to the problems on which state-of-the-art models for object detection and classification are usually being evaluated. In the first step of tackling these challenges, we used the well-known classification backbone from the EfficientNet family and compared Softmax loss with Additive Angular Margin Loss (ArcFace) [2]. ArcFace is used to obtain highly discriminative features of visually similar products. First results on sub-dataset with 25471 images and 156 classes of pâtés showed ArcFace, with an accuracy of 95.84%, outperforms Softmax loss model, which achieves 95.19%. In future work, we will expand our datasets, develop and evaluate new deep learning models, and build contextual information from surrounding products to improve classification accuracy.

Out-of-stock (OOS) is a problem that all stores face as a part of their daily business. Corsten and Gruen [3] conducted research on the extent, the causes, and

the efforts to address OOS situations. They measured OOS rate as the percentage of unique products which are not present on store shelves at a particular moment in time. Results showed that the average worldwide OOS rate is 8.3%, while the estimated loss of sales which those missing products generate is 3.9%. Therefore, implementing a fast and accurate OOS detection system should have a positive impact on sales results. We manually annotated OOS situations in our in-house dataset that consists of several hundred high-resolution images of supermarket shelves. Exploratory data analysis showed that OOS situations are rather small, covering around 1% of the whole image and that there are 2.2 OOS situations per image. We trained YOLOv5 detection model on a subset of data which contains shelves images with displayed bottles and initial results are: 0.56 precision, 0.59 recall and 0.58 mAP. In future experiments, we will use more data and different deep learning models in order to detect OOS situations with higher accuracy.

This project has several other problems which need to be addressed such as price tag detection, out-of-distribution problem, recognition of different very similar stock keeping units (SKUs) e.g., 0.5 liter bottle of some beverage and 2 liters bottle of that same beverage, etc. Solving these problems should help us reach both better detection performance and higher classification results.

References:

1 Wei, Y., Tran, S.N., Xu, S., Kang, B.H., & Springer, M. (2020). Deep Learning for Retail Product Recognition: Challenges and Techniques. Computational Intelligence and Neuroscience (pp. 1-23).

2 Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).

3 Corsten, D., Gruen, T. (2005). On shelf availability: An examination of the extent, the causes, and the efforts to address retail out-of-stocks. In Consumer Driven Electronic Transformation (pp. 131-149).

Feature Matching Methods for surround-View Camera Systems

D. Khowaja, M. Subašić, S. Lončarić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Large vehicles like cranes, fork lifts, dump trucks, excavators, trucks etc. have large blind spots areas. Blind spots are the regions in the surroundings of all vehicles and mobile heavy machinery where the operator has no visibility. These blind spots can be large enough to hide from view, other vehicles, cyclist or even commuters and bystanders. Several accidents are caused due to blind spots around these heavy vehicles. These accidents mostly take place in complicated

environments such as while executing turns, intersections or roundabouts after driver fail to detect other road users in the blind spots which are in the region very close or alongside their vehicles. Around 400 people in Europe are estimated to be killed every year in such situations, most of them being vulnerable road users such as cyclists, bike riders and commuters. The European Union has introduced a Rule requiring new heavy vehicles to have a mirror or camera device covering the blind spot area in the vicinity of the vehicle to avoid accidents. New vehicles are therefore fitted with a lot of mirrors, but the vehicle's entire surrounding is still not protected.

Computer Vision - based Advanced Driver Assistance Systems are known to increase driver safety by providing visual information of the vehicle surroundings. Despite being an integral part of most cars, ADAS systems are being added to heavy machinery only recently. ADAS systems using multiple cameras can be used for surround view visualization of complex heavy machinery. Finding image correspondence is a basic computer vision problem and a gateway to many applications such as 3D reconstruction, video processing, image recovery, image stitching and object recognition. For instance, one of the primary requirements for developing bird's view is proper stitching of images which can be achieved if the features in images are matched correctly. These matches can be used to align the two images and create a panorama.The images from multiple camera are matched, stitched and optimized to deliver a singular surround view which makes maneuvers easier and safer. Feature matching algorithm act as the basis for several computer vision tasks.

Three main steps involved in feature matching are feature detection, feature description and matching of descriptors. A typical feature matching algorithm starts by detecting keypoints in two images to be matched. Features can vary from a single pixel to corners and edges and can be as large as objects in the image[2]. Feature detection is the process of identifying these Interest Points in an image. Usually, the output from a feature detector is a number of individual locations in an image, called points of interest or keypoints. Then, descriptors for those keypoints are calculated using feature descriptors. Feature description defines each feature by giving it a unique identity which facilitates efficient recognition for matching. The neighborhood around each keypoint is described in a way that is invariant to illumination, translation, rotation and scale. Typically the output from a feature descriptor is a feature vector descriptor surrounding each interest point. Several feature detector algorithms are individually available for performing feature detection while some are available with their respective feature descriptor algorithms. The individual feature detectors can be combined with appropriate feature descriptors. To determine correspondence between descriptors in two images, feature-matching is performed. Similar features are

detected by comparing the descriptors. The ouput from a feature matcher is a set of points (Yi , Zi) (Yi' , Zi'), where (Yi , Zi) is a feature in first image and (Yi' , Zi') is its matched feature in second image. Various Feature matching approaches can be implemented for matching features such as Nearest Neighbor; Threshold based matching ; Nearest Neighbor Distance Ratio etc. depending on there strengths and limitations. Generally the selection of feature detectors and descriptors greatly affects the matching performance. Therefore, feature detectors and descriptors suitable for images contents should be used in applications. For instance, with images of bacteria cells, a blob detector shall be used instead of corner detector. For images with man-made structures like aerial view of a city, a corner detector shall be used. Moreover, it is very important for a detector or descriptor to address the image degradation[2].

Recently, many new approaches were proposed and shown to perform better than previous alternatives on standard benchmarks. However, whether it is useful to integrate them into the standard pipeline is less-researched. In this research, we conduct an extensive evaluation of recently proposed algorithms by integrating them into the well-established image matching pipeline to investigate whether they can improve the overall matching performance. To address this, we create a large-scale dataset comprising of around 12000 sythetic images that cover an extensive variety of scene types, containing image pairs exhibiting a range of dramatic variations in illumination, geometry and appearance. The ground truth matches for all image pairs are produced using Epipolar geometry constraints, which is a basic concept in image geometry which reveals the relationship between two views. This concept is utilized to get highly accurate ground truth matches in an image pair. Based on this ground truth, state-of-the-art matchers and feature extraction methods are exhaustively evaluated and analyzed.The experiments are conducted using strictly defined evaluation metrics, and the implying results provide insight into which algorithms perform better in which scenarios.

We set a standard pipeline as the reference. Precisely, we use DoG (difference of Gaussians) as detector and SIFT (Scale-invariant feature transform)[1] as descriptor to create primary correspondences among images by the nearest-neighbor search, then eliminate bad correspondences using Lowe's ratio test, and finally remove outliers using GMS (Grid-based Motion Statistics). For each evaluated algorithm, we incorporate it into the baseline system by replacing its counterpart, and use the overall performance for comparison. we evaluate GMS, a prunning method, with different combinations of deep learning and hand-crafted detectors and descriptors. Finally we remove outliers from matched points that are more than three scaled median absolute deviations. The different combinations of deep learning based and hand-crafted local features that are evaluated include[3]:

- HessianAffNet(detector), HardNet++(descriptors) and GMS.

- DoG(detector), HardNet++(descriptors) and GMS.
- DoG(detector), SIFT(descriptors) and GMS.
- SIFT(detector), SIFT(descriptors) and GMS.
- HessianAffNet(detector), SIFT(descriptors) and GMS.
- ORB(detector,descriptors) and GMS.
- D2Net(detector,descriptors) and GMS.
- Contextdesc and GMS.
- SURF(detector,descriptors) and GMS.

We use Matlab functions for implementation of SURF, VLFeat library for implementing SIFT detector, descriptor and DoG detector, and the threshold is 0.8 for ratio test. Other codes are from authors' publicly available implementations, where we use the pre-trained models published by authors for deep learning based methods. The evaluation results are presented using curves where matching performance is evaluated using Error versus angle between cameras and ratio of number of Matches from matcher to correct number of matches from ground truth.Moreover, ratio of Number of Matches from matcher to number of keypoints from detector and number of matches from each combination versus angle between cameras are the metrics used for analyzing the results. The preliminary results demonstrate following two combinations for achiving best matching results. 1: DoG(detector), HardNet++(descriptors) and GMS. 2: HessianAffNet(detector), HardNet++(descriptors) and GMS. The in-depth analyses based on the reported results can be used as general guidelines for designing practical image matching systems.

References:

1 Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal on Computer Vision (IJCV) 60 (2004) 91–110.

2 Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. International Journal on Computer Vision (IJCV), 65(1-2):43–72, 2005

3 Bian,J.-W.,Wu,Y.-H.,Zhao,J.,Liu,Y.,Zhang,L.,Cheng,M.-M.,etal."An evaluation of feature matchers for fundamental matrix estimation". In British machine vision conference (BMVC) 2019.

A Topic Coverage Approach to Evaluation of Topic Models

D. Korenčić, S. Ristov, J. Repar, J. Šnajder

Ruđer Bošković Institute, Zagreb, Croatia, and Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Topic models are popular unsupervised models capable of extracting topics from large text collections. Topics, represented as a weighted list of words and documents, are expected to be interpretable as concepts. Due to the inherent stochasticity in learning algorithms and the models' inductive bias, topic models are susceptible to errors that manifest as incoherent topics containing unrelated or random words and documents. Methods of topic model evaluation offer measures of model quality that can serve both as a tool for model analysis and as a guide for model construction. Examples of such methods include measures of semantic coherence of topics and measures of model stability. The coverage approach to topic model evaluation relies on a set of pre-compiled reference topics and the measures of coverage. Reference topics are topics that the models are expected to discover in a certain application scenario, while coverage measures quantify the match between generated model topics and reference topics. The coverage approach enables large-scale automatic evaluation of both topic models and other measures of model quality.

We propose a coverage approach rooted in the use case of topic discovery - interpretation of many model topics performed to discover useful concepts occurring in a text collection. In our experiments, we use two datasets of reference topics derived from topic discovery experiments performed on news and biological text. We propose two new measures of coverage - a supervised measure that approximates human intuition of topic matching, and an automatic unsupervised measure. The supervised matching model is simple, grounded in high-quality annotations of topic similarity, and achieves performance close to the human agreement. The output of the supervised measure is the proportion of the reference topics that match, i.e., are covered by, at least one model topic. The unsupervised measure uses a measure of topic distance and a distance threshold to approximate the matching decision. The threshold-based coverage scores are integrated over a range of thresholds to derive the final coverage score. We show that the unsupervised measure, the first such measure of coverage, approximates supervised coverage scores extremely well. The method can be used both for coverage-based model selection and for graphical analysis of model performance.

To demonstrate the merits of the coverage approach, we conduct a series of experiments on a dataset of political news text and a dataset of biological texts about microorganisms. News reference topics correspond to persons, organizations, events, and abstract concepts, while the biological topics correspond to

phenotypes (organism characteristics). In the first set of experiments, we evaluate the coverage of reference topics obtained by topic models of varying types and sizes. We measure the overall coverage, and the coverage of topics categorized by size and semantic class. The experiments show that the NMF model based on matrix factorization achieves competitive performance across both datasets. Additionally, we empirically demonstrate the difference between topic models of different sizes - models with large topic capacity detect reference topics of all sizes, while the smaller models are only able to detect large reference topics (topics occurring in many documents). These findings yield practical recommendations for the use of topic models in topic discovery - large models and NMF models are expected to achieve good coverage.

The second set of experiments applies the coverage methods to evaluate two established model evaluation approaches - topic coherence and model stability. The coherence experiments demonstrate that the state-of-art coherence measures have neither high nor consistent correlation with coverage. This finding is in line with the previous experiments and reaffirms the need for re-examination of the widely used coherence measures. The stability experiments demonstrate that no correlation exists between coverage and stability, which suggests that stability alone is not a reliable guide for the selection of high-quality models, as is implied in several earlier experiments. Additionally, we show that the proposed coverage measures can be successfully adapted for the calculation of model stability.

Measures of topic coherence aim to quantify the vaguely defined quality of topics' semantic coherence, sometimes also defined as the correspondence of a topic to some concept. Coherence measures, based on an intuitive notion and easy to compute, have since their introduction become a de-facto standard for semantic evaluation of topic models. However, several recent experiments demonstrated that coherence measures achieve either inconsistent or weak correlation with quality scores based on human inspection of topics. These results, in line with our findings related to coherence, represent arguments for the pressing need to devise new automatic measures of model quality. Namely, quality measures based on human inspection and evaluation of topics are slow and impractical to deploy. The coverage approach cannot replace the measures of coherence, since the coverage experiments in essence represent a simulation of a fixed topic modeling scenario defined by a collection of texts and a set of reference topics. The strength of the coverage approach is its potential to guide the analysis and design of both topic models and coherence measures. Namely, the coverage experiments can be automatized, and the measures of coverage are well defined since they quantify correspondence to a fixed set of reference topics. Therefore, if the number of pre-compiled coverage datasets is increased, this would enable large-scale semantic evaluation of both topic models and measures of model quality.

Our research on topic coverage resulted in new and practical measures of coverage, including the first unsupervised measure of coverage. It also yielded insights into topic models and practical recommendations for their use, and findings about other methods of topic model evaluation - topic coherence and model stability. There exist many directions for future work on topic coverage. The measures we propose could be improved by making the supervised measure more accurate and quicker to construct by way of using the active learning techniques. The unsupervised measure should be made interpretable to extend its use beyond producing model rankings. A graphical tool that would enable easy construction and management of reference topics would greatly facilitate future experiments. Future coverage experiments should be conducted on new datasets and with different model architectures in order to obtain more robust findings and recommendations. The field of computational social sciences, where topic models are often applied as tools for quantitative research, could benefit from topic coverage. Namely, such investigations often expect the models to pinpoint specific theoretical concepts such as agenda issues and media frames, and the coverage experiments with reference topics that correspond to such concepts could evaluate topic models for specific use cases. In general, we believe that future research on topic coverage has the potential to provide new and more precise data on the performance and applicability of both existing and future topic models.

References:

1 D. Korenčić, S. Ristov, J. Repar, and J. Šnajder, "A topic coverage approach to evaluation of topic models," IEEE Access, vol. 9, 2021.

2 D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010, pp. 100–108.

3 A. Hoyle, P. Goel, D. Peskov, A. Hian-Cheong, J. Boyd-Graber, and P. Resnik, "Is automated topic model evaluation broken?: The incoherence of coherence," arXiv preprint arXiv:2107.02173, 2021.

Multidisciplinary Data Intensive Applications

Data-driven Approaches for Mobility: Collection, Contextual Enrichment and Analytics of Automotive Data

H. Vdović, D. Pevec, J. Babić, V. Podobnik

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

In the last several years, companies are increasingly using data processing and analysis to make informed engineering and business decisions. This can perhaps best be seen in the increase of revenue from big data and business analytics, which is projected to grow to $274 billion in 2022, while it was worth 122 billion in 2015. Software is the key driver for creating insights from previously collected data. The decision to use software to aid data analysis in a particular project may be influenced by a number of factors, such as the nature of the data and the complexity of the problem at hand.

Recently, there has been a significant increase in the amount of data collected in virtually all industries operating today, and the automotive industry is not different. This is mainly due to the rising adoption of connected cars which are capable of communicating with the Internet, and can therefore be utilized for data collection. Vehicle generated data is valuable for use cases such as fleet management, predictive maintenance, driver behaviour profiling and it is especially useful in applications dealing with transportation sustainability, as the vehicle's energy efficiency and energy consumption can be measured and monitored using automotive data. A way of providing even better insights from the collected automotive data is to enrich it with context. Contextual enrichment encompasses the identification of additional information sources which can be used to enrich the available data and contribute to its interpretation. Location and weather condition information can facilitate the generation of new insights when analysing the existing automotive data. With the addition of other location-based information sources such as traffic data, researchers can come to even more useful conclusions about the vehicle and driver performance.

The major obstacle for performing research in this area is that there is no open-source contextually enriched automotive data set available. We aim to overcome this obstacle by providing a solution for collection of such a data set in the form of a framework for collection and contextual enrichment of automotive data. The proposed framework was modelled and developed in a way that satisfies the main research questions which were stated in the introduction of the paper. The data collection process was done by utilizing smartphones as the data collection tool to cover the widest possible range of diverse vehicles. Also, two automotive communication technologies - OBD-II and CAN are supported

by the framework to further maximize the spectrum of vehicles from which the data can be collected. Additionally, to facilitate the development of CAN bus collection modules, a CAN bus simulator was modelled, developed and validated using multiple mathematical similarity measures. Smartphones also allow the contextual enrichment of automotive data using its built-in sensors and Internet connection. The challenge of contextually enriched automotive data storage was solved by implementing a custom distributed data storage platform. The data platform was constructed in a manner that addresses the requirements of reliability, scalability, and fault tolerance, as well as requirements imposed by the big data complexity, and was validated on several electromobility use cases. The final part of the framework is the application programming interface which serves as a middleware solution between the end-users and the data storage platform and is used for accessing and analysing the stored data [1]. Although the data collection process and data model used by the API were explained to provide a better understanding of the API's functionalities, the API is agnostic to the data collection process. Therefore, the API can also be used with a different storage solution behind it by changing the database connection library utilized by the API. In any case, the described API can be used as a blueprint for the design and implementation of an API for advanced analytics intended to be used with automotive data.

The final presented contribution is an open source contextually enriched automotive data set collected as a part of the data collection experiment using the established and implemented framework. The data set consists of 287,882 data points collected during 212 trips made by 9 drivers, resulting in a total of 94 hours, 25 minutes and 44 seconds of recorded driving time. This data set has been made publicly available for researchers to access and use.

The data set was analysed in several different use cases. We demonstrated how such a data set can be used in the field of transportation sustainability to evaluate the eco-efficient driving patterns of a group of drivers. For this reason, a metric named eco index was established which incorporates for eco-efficient driving advice and the drivers who participated in the data collection experiments were ranked accordingly. The established evaluation process is mathematically accurate and can be useful for applications such as vehicle fleet management, where cost reduction can be achieved through improved fuel efficiency [2]. Additionally, driving patterns were evaluated on the level of a single trip, and it was demonstrated how the trips can be divided into groups according to the similarity of their properties [3]. Three groups were identified, which represent an interpretive classification of the trip. The results of the analysis clearly identified differentiating characteristics of each group. That said, Group 1 represents fast rides, long distances without major stops. Group 2 and 3 are both at lower speeds. Group

2 is characterised by a desire for accelerated movement and tendency to drive above the speed of the rest of the traffic. Group 3, on the other hand, includes trips with a smoother driving style that blends in with the surrounding traffic. The last data analytics use case demonstrated was a driving route comparison measurement which was established and implemented in a software system for interactive visual analysis of the collected data set. The system allows a user to detect the best route for the selected start and end location, analyse the driving style of each driver, compare driver behaviour in different weather conditions and provide insight on the driver's speed compared to the rest of the traffic on each trip which can be used to ultimately improve their driving style.

Acknowledgement: The authors acknowledge the support of the European Regional Development Fund under grants KK.01.2.1.01.0020 (RASCO-FER-SMART-EV), KK.01.2.1.02.0071 (MUNIVO) and KK.01.1.1.01.0009 (DATACROSS).

References:

1 H. Vdovic, J. Babic, and V. Podobnik. "An Application Programming Interface for Advanced Analytics of Contextually Enriched Automotive Data." 2021 16th International Conference on Telecommunications (ConTEL). IEEE, 2021.

2 H. Vdovic, J. Babic, and V. Podobnik, "Eco-efficient driving pattern evaluation for sustainable road transport based on contextually enriched automotive data." Journal of Cleaner Production 311 (2021): 127564.

3 I. Gace, D. Pevec, H. Vdovic, J. Babic, and V. Podobnik. "Driving style Categorisation based on Unsupervised Learning: a Step towards Sustainable Transportation." 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE, 2021.

Influence of Chimeric Sequences on Metagenome Assembly

M. Huang, M. Šikić

*Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), and University of Zagreb, Faculty of Electrical Engineering and Computing*

While there has been many advances in genome sequencing, obtaining a large amount of high quality data can be expensive and time-consuming. In addition, artifacts like chimeric sequences (or split sequences) can form when two or more are incorrectly joined together in the process of genome sequencing. Regardless of technology used, these chimeric sequences can still propose a challenge for downstream analysis of genomes, either providing wrong information about the order of sequences or force us to forfeit these information to maintain the integrity of data.

In the research of de novo assembly, assemblers overlap short nucleotide sequences to form long sequences without the aid of a reference genome. Chimeric sequences can hinder this process by providing incorrect overlap information. Existing assemblers try to process these chimeric sequences by filtering or splitting them into separate correct fragments. However, detecting these chimeric sequence can be difficult when data quality and/or quantity is low. For instance, with low quality data, differentiating between errors in bases and chimeric sequences can be tricky. There is a risk of discarding important genomic information instead. Without a reference genome, sequences can also be wrongly identified as chimeric. The focus of this work is to investigate the influence chimeric sequences have on genome assembly, with existing assemblers and their chimeric sequence processing features, by analyzing their resultant assembly with and without chimeric sequences.

To facilitate the detection of chimeric sequences, Ratlesnake is used. It is a sequence classification tool that uses a given reference genome to find breaking points in chimeric sequences, regions of repetitive sequences and containment. A dataset with simulated chimeric sequences is used to optimize the accuracy, precision and recall rate of Ratlesnake's chimeric detection before it was used to evaluate the dataset in this work.

The dataset used for evaluation is a mock community of 10 microbial species (ZymoBIOMICS Microbial Community Standards) sequenced with Oxford Nanopore GridION. MetaQuast is used as the evaluation criteria. The dataset is analyzed and compared in two forms: (1) whole dataset and (2) dataset with chimeric sequences filtered with various chimeric detection tools.

First, Ratlesnake is used to detect chimeric sequences in the dataset. The data is then processed and prepared in the two forms where identified chimeric sequences are either kept or removed by Ratlesnake. These processed datasets are then assembled with Miniasm [1], Raven [2] and Flye [3]. MetaQuast is used to evaluate the assemblies before and after chimeric sequences are removed from the dataset. The same set of evaluation is conducted for the various assemblers as well. Through this pipeline, it is observed that the number of misassemblies by most assemblers decreased. For some microbial genomes, there was a significant increase in NGA50 while the assemblies of other genomes was maintained.

As this work went through changes in versions and thus, quality, of the dataset used, assemblies got drastically better as well. This effect was similarly observed with new releases of Minimap2, which MetaQuast relied heavily on. While data and evaluation tool quality played a big role in the completeness of assemblies, the positive effect of chimeric sequence detection and removal remained consistent.

References:

1 Heng Li, Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences, Bioinformatics, Volume 32, Issue 14, 15 July 2016, Pages 2103–2110

2 Vaser, R., Šikić, M. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1, 332–336 (2021)

3 Kolmogorov, M., Bickhart, D.M., Behsaz, B. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110 (2020)

Motorway Bottleneck Detection Using Speed Transition Matrices

L. Tišljarić, F. Vrbanić, E. Ivanjko, T. Carić

Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia

Motorway bottlenecks are commonly occurring in developed countries with increased commercial activities. As opposed to the traffic jam, the source of the bottleneck is often a result of a specific traffic situation like traffic accidents, sudden breaks, or bad road design. They could be described as a sudden capacity drop that creates congestion on the part of the motorway that propagates in the downstream flow. More formally, free flow (normal traffic flow) could be described with equation $q_{in} \approx q_{out}$, where q_{in} represents upstream flow and q_{out} represents downstream traffic flow on the observed part of the motorway. The bottleneck manifested through the decrease of the downstream traffic flow due to the congestion. There are two types of bottlenecks: recurrent and non-recurrent. The recurrent bottlenecks are caused due to the influence of daily commuters and transit traffic. At the same time, non-recurrent are often the result of traffic anomalies like traffic accidents or some unexpected events [1].

This paper aims to present a method for motorway bottleneck detection by computing the probability of bottleneck occurrence using Speed Transition Matrices (STMs). The method is divided into three parts: (i) preprocessing of synthetic Global Navigation Satellite System (GNSS) dataset extracted from simulation, (ii) STM computation, and (iii) bottleneck probability estimation. The first step includes preprocessing the raw GNSS data to extract the routes of every vehicle, dividing the route into the transition between motorway segments, and computation of the mean harmonic speeds for every vehicle on motorway segments. The second step includes STM computation by counting the speed transitions and presenting them by speed transition probability distributions, explained in the next paragraph. The third step incorporates STMs and the Center of Mass (CoM) computation as an input to the fuzzy-based system for the bottleneck probability estimation. The contributions of this paper can be

summarized as: (i) proposed method for the motorway bottleneck detection using the STMs, (ii) evaluation of the method on synthetic dataset under different traffic scenarios.

As a traffic data modeling method, the STM is proposed [2]. The STM captures the vehicle's speed at the movement between two consecutive motorway segments called transition. It is used to represent the speed probability change, and therefore it represents the speed probability distribution at one transition in one time interval. Captured speed is relative to the speed limit on the motorway road segments that are observed. In this paper, 5% is chosen as the discretization period, and 100% is the maximal possible relative harmonic speed, which resulted in matrix dimensions of 20×20. Therefore, every row represents the speed change on the origin edge e_i of 5%, and every column represents the speed change on the destination edge e_j of 5%. The value in the cells of the STM p_{mn} represent the probability of the relative speed change between speed value from m to n.

There are five examples of the characteristic STMs on the motorway: (i) vehicles on the observed transitions had, both origin and destination speeds, close to the 100% of the speed limit, where traffic state can be defined as the free-flow, (ii) more unstable traffic state because vehicles are traveling with speeds close to 60% of the speed limit, (iii) beginning of the bottleneck is characterized with the transitions from high speed values to low values, (iv) end of the bottleneck where vehicles are transitioning from congested traffic flow to free-flow, and (v) congested traffic state is characterized by very low speeds on both origin and destination segments of one transition.

To extract the most probable transition, the CoM for each computed STM is computed. The method was adopted from [2], which relies on the computation of the expected speed values. When the CoM is computed, features suitable for quantification of the bottleneck probability were selected. Two features were extracted from the computed CoM: (i) distance from the origin of the STM labeled as d_S and (ii) distance from the diagonal of the STM labeled as d_D. The first feature, d_S, is important for estimation of the traffic state on the observed transition. If the CoM is closer to the origin of the STM, it represents traffic with low speeds on both origin and destination segments, and as d_S grows, origin and destination speed grows, which leads to free-flow conditions. On the other hand, larger d_D values will show sudden breaks area and intense acceleration area of the STM.

As features d_S and d_D can be represented as linguistic variables, the Fuzzy Inference System (FIS) is used to detect the probability of the bottleneck occurrence [3]. In its initial form, FIS consists of two input variables d_D and d_S with corresponding output p_b that represents bottleneck probability. All variables are represented with range $[0, 1]$, relative to their maximal values. The maximal

value of the d_S is is the length of the STM diagonal that can be computed as $20\sqrt{2}$, while the maximal value of d_D can be computed as $d_S/2$.

The proposed bottleneck probability estimation method was evaluated on the synthetic dataset extracted from the Simulation of Urban MObility (SUMO) traffic simulation framework. The method was evaluated on four different possible motorway congestion scenarios that originate from different congestion sources: traffic accident, short recurrent congestion from on-ramp flow, long recurrent congestion from on-ramp flow, and moving bottleneck originated from slow, heavy-duty vehicles on the motorway. The method was evaluated and compared to the measured speed values, with the initial results indicating the usability of the method in every proposed scenario.

Future work will include validation of the results by comparing baseline results of the three most crucial traffic parameters: speed, density, and traffic flow. An envisaged future research direction will include the usage of a convolutional neural network for automatic extraction of the STM features, as they can be used as traffic images.

References:

1 L. Tišljarić, S. Fernandes, T. Carić, J. Gama, "Spatiotemporal Traffic Anomaly Detection on Urban Road Network Using Tensor Decomposition Method," International Conference on Discovery Science, Sep 2020., pp. 674-688

2 L. Tišljarić, T. Carić, B. Abramović, T. Fratrović, "Traffic State Estimation and Classification on Citywide Scale Using Speed Transition Matrices," Sustainability, vol. 12, no. 18, pp. 7278, Sep 2020.

3 L. Tišljarić, Z. Kavran, E. Ivanjko, T. Carić, "Fuzzy Inference System for Congestion Index Estimation Based on Speed Probability Distributions," Transport Research Procedia, vol. 55, pp. 1389-1397, Jan 2021.

Agent-based Urban Traffic Control

E. Ivanjko, G. Martin, D. Čakija, K. Kušić, M. Miletić, F. Vrbanić, Majstorović

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

Intelligent Transportation Systems (ITS) offer a solution for alleviating congestion in urban areas based on its traffic control service. It is part of the build + ITS approach where new infrastructure with different ITS services is applied to improve traffic throughput, reduce vehicle emissions, increase the Level of Service (LoS), and improve the usage of existing vehicles with the Mobility as a Service (MaaS) concept. The latter enables the cities to offer mobility users concepts of car-sharing, ride-sharing, and tailored multi-modal travel patterns.

The city or urban road infrastructure can be divided into urban motorways and networks of signalized intersections. Traffic control approaches related to Variable Speed Limit (VSL) control, ramp metering, lane change recommendation, and Adaptive Traffic Signal Control (ATSC) are applied to manage the urban road infrastructure. Recently, approaches based on Machine Learning (ML) are being researched and deployed. These approaches are usually learned by running multiple simulations using microscopic traffic simulators prior to real-world application. The simulations need to be as realistic as possible to capture the real-world traffic behavior, including parameters like speed limits and right of way. Microscopic simulations validate or refute the (dis)advantages of the proposed traffic control approaches. Thus, multiple scenarios need to be created to obtain and analyze traffic behavior. To create relevant scenarios, k-means, Principal Component Analysis (PCA), and Self-Organizing Maps (SOM) can be applied to analyse the induction loop detector data. K-means clustering is a widely used technique for data analysis, where a cluster represents a collection of aggregated data measurements that share certain similarities. PCA is defined as an orthogonal linear transformation that transforms the data into a new coordinate system. The largest variance by a scalar projection of the data lies on the first coordinate (called the first principal component), the second-largest variance on the second coordinate, and so on. These principal components represent the data patterns of the main data set. Dimensionality reduction makes it easier to visualize and process high-dimensional data sets while preserving as much of the eigenvalue as possible in the data set. The SOM neural network allows the visualization of high-dimensional data and neural distances between neurons. Learning a SOM aims to create a model that places similar input data (vectors) based on the calculated Euclidean distance to the closest neuron, called the Best Matching Unit (BMU). Analyzing the clustering results on the number of relevant scenarios is necessary to capture traffic behavior accurately. Those created relevant scenarios can be used for teaching of ML based traffic control systems. One of the ML approaches is the Advanced Deep Reinforcement Learning (DRL) framework based on Actor-Critic architecture, and it can be applied for more complex and holistic strategies such as differential VSL control set in cooperation with ramp metering. The Deep Deterministic Policy Gradient (DDPG) algorithm as the DRL method is currently the most prominent approach in scientific studies for motorway control since it can provide continuous actions. Depending on the model structure used in the Actor-Critic module design, it is possible to process inputs as an image-alike representation of motorway traffic or as the vectorized representation of macroscopic traffic parameters. The mentioned control approach is made at the level of a single DRL agent with the main goal to maximize average speed on the controlled motorway section. Moreover, current research addresses the use of different DRL-based VSL strategies with respect to their configuration

and control comprehensiveness at one characteristic traffic scenario for urban motorways. The expansion of the DDPG based DRL approach towards multi-agent architecture is investigated in the form of agent-based workers who are learning in parallel by sharing knowledge learned through different traffic scenarios/environments. The multi-agent DDPG approach can be made through the Advantage Actor-Critic (A2C) or Asynchronous Advantage Actor-Critic (A3C) architecture, which will consequentially introduce more comprehensive and robust control over stochastic traffic flow fluctuations on the controlled motorway section. Most studies in Reinforcement Learning (RL) based VSL (RL-VSL) control are based on a single explicit objective or multiple objectives implemented as a single control policy. However, large-scale control systems usually consider various, often conflicting objectives with different time and space scales (simultaneous optimization of VSL and ramp metering) or different priorities (safety versus throughput or higher traveling speed). Ideally, the VSL application area should be split into several smaller VSL sections upstream of the bottleneck area to adjust the speed limit for upcoming traffic flow gradually. The requirements mentioned above can be modeled and solved by using Multi-Agent RL-based (MARL) control approaches. Using MARL techniques, the speed limits can be controlled in multiple strategically placed VSL zones simultaneously, implementing multiple policies (with different objectives) among agents simultaneously, thus, enabling heterogeneity between RL agents in VSL control [1]. In the case of intersection control in urban areas, it is evident that daily patterns of morning and afternoon peak periods cause congestions. As building additional infrastructure is not an option, the research focus shifted towards better use of existing infrastructure. ITS technologies such as ATSC can improve the usage of existing urban road infrastructure. By measuring the current traffic state in real-time, ATSC systems can change the corresponding signal program to accommodate the changes in vehicle flow fluctuations better. In most cases, the queue lengths on each intersection approach are measured to create a feature state space. By applying RL techniques in ATSC systems, the agent can learn the optimal control policy during operation. The principal problem with such agents is the large state space making the learning process too slow for a technical system. By reducing the complexity of the state space, the speed of the learning process can be increased without any impact on the final result. SOM and Growing Neural Gas (GNG) are techniques that can reduce the dimensionality of the state space, with the added benefit that GNG state representation can adapt while the agent is still learning. This approach allows the agent to learn both the state representation and action selection simultaneously [2]. The application and impact of a new traffic control approach tailored to traffic flow that contains Autonomous Vehicles (AVs) and CAVs on traffic safety, flow, speed, fuel consumption, and emissions have to be analyzed. One commonly used approach for urban motor-

ways is the VSL control algorithm that increases the LoS by reducing the speed of vehicles incoming to a bottleneck area in classical traffic flows containing only Human Driven Vehicles (HDVs). The feedback-based VSL can be improved with RL. Q-learning with discretized states and actions can improve VSL operation. CAVs can be utilized as posted speed actuators using I2V communication through Road-Side Unit (RSU). Therefore, classical VMS is replaced, and CAVs must be equipped with an On-Board Unit to obtain the speed limit information [3]. It is also necessary to analyze different CAV penetration rates to understand the implications of introducing CAVs in mixed traffic flows. Current results support the conclusion that the need for separate traffic control systems reduces with the increase of CAV penetration rate. Thus, CAVs can resolve some of the traffic congestion problems by better vehicle driving and traffic data sharing. As a result, a traffic control system can have the ability to improve its control quality during operation. This is crucial when traffic patterns are changing under the influence of seasonality or new ways of using the transport infrastructure. The MaaS concept and influence of the COVID-19 pandemic are examples of creating an urban environment with rapidly changing mobility patterns. Contrary to classic traffic control system, newly developed ML-based traffic control system can adapt dynamically according to the existing of modified criteria functions. Additionally, learning structures can easily include additional measurement inputs being available with the dawn of mixed traffic flows containing HDVs and (C)AVs. To ensure the successful deployment of such systems, open questions related to learning convergence and optimal representation of the state-action space have yet to be tackled and resolved.

References:

1 K. Kušić, E. Ivanjko, F. Vrbanić, M. Gregurić, I. Dusparic, "Dynamic Variable Speed Limit Zones Allocation Using Distributed Multi-Agent Reinforcement Learning," In Proceedings of the 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC), 2021, pp. 3238-3245, doi: 10.1109/ITSC48978.2021.9564739

2 M. Miletić, E. Ivanjko, S. Mandžuka and D. K. Nečoska, "Combining Neural Gas and Reinforcement Learning for Adaptive Traffic Signal Control," 2021 International Symposium ELMAR, 2021, pp. 179-182, doi: 10.1109/ELMAR52657.2021.9550948

3 Vrbanić, F.; Ivanjko, E.; Kušić, K.; Čakija, D. Variable Speed Limit and Ramp Metering for Mixed Traffic Flows: A Review and Open Questions. Appl. Sci. 2021, 11, 2574. doi: 10.3390/app11062574

Financial Time Series Labeling Evaluation

T. Kovačević, F. Šarić, L. Mrčela, S. Begušić, Z. Kostanjčar

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

The widespread application of machine learning models in various scientific disciplines has led to the increasing use of such models for predicting price movements in financial markets. To show that these models are superior to standard statistical methods, emphasis has been placed on the models themselves and various combinations of input features. However, as pointed out in [1], there is no canonical way of labeling financial time series.

Generally, the problem of predicting the dynamics of financial time series can be reduced to regression and classification. The regression approach predicts returns, raw prices, or volatility for the fixed future horizon. The classification approach usually predicts thresholded future returns [2]. However, there are many other methods to obtain class labels, such as non-causal technical indicators, local trend detection algorithms, etc.

It is well known that financial time series are noisy, they only possess poor, hardly identifiable information, and the data sample is relatively small; therefore, any chosen labeling algorithm may produce noisy labels. In numerous studies of regression target prediction, filtering out financial time series has been shown to benefit model performance (for both regression metrics and financial performance metrics of the strategy that relies on model output). However, the impact of filtering financial time series prior to classification labeling has not been investigated yet [3].

In this paper, we propose an evaluation method for different financial time-series labeling algorithms. The idea of the evaluation method is to see how a strategy based on supervised learning would perform considering the model generalization error. To do this, the generalization error is simulated by corrupting the labels initially obtained by a labeling algorithm and then backtesting a strategy with the corrupted labels. The evaluation method consists of four main components: machine learning-based trading strategy (i.e., take a long position if the model outputs a positive class), labeling algorithm that generates classes (i.e., if the future price is greater than the price at a given moment, the market movement is upward and therefore the class should be positive), corruption function that simulates the generalization error of a machine learning model (i.e., corrupted entry and exit moments of positions initially obtained by a labeling algorithm), and a financial performance measure to evaluate a trading strategy based on machine learning model output (i.e., cumulative return, Sharpe ratio, etc.).

Moreover, we propose a novel algorithm for labeling financial time series, which we call Oracle labeling. Given the historical prices of some financial assets and the fixed transaction costs for opening and closing long and short positions, the algorithm determines the optimal moments for opening and closing long and short positions, which yield theoretically maximal cumulative return.

We evaluate the newly proposed Oracle labeling algorithm and the N-step-ahead labeling algorithm, the most commonly used labeling method in the reviewed literature, using the proposed financial time series labeling evaluation method. We also investigate the impact of filtering financial time series using the Hodrick-Prescott filter prior to labeling price series.

Empirical analysis of both algorithms suggests that the Oracle labeling algorithm is generally more robust to corruption in terms of financial performance measures than the N-step-ahead method for several types of financial assets. We also show that filtering financial time series prior to labeling generally increases label robustness for both labeling algorithms.

References:

1 P. N. Kolm, J. Turiel, and N. Westray, "Deep Order Flow Imbalance: Extracting Alpha at Multiple Horizons from the Limit Order Book," SSRN Electron. J., Aug. 2021.
2 M. M. Ĺópez de Prado, Advances in financial machine learning.
3 M. Ouahilal, M. El Mohajir, M. Chahhou, and B. E. El Mohajir, "A novel hybrid model based on Hodrick–Prescott filter and support vector regression algorithm for optimizing stock market price prediction," J. Big Data, vol. 4, no. 1, pp. 1–22, Dec. 2017.

The Applicability of Functional Clustering in Analyzing Historical Floods of the Sava River in Zagreb

M. Lacko, K. Potočki, D. Pintar, L. Humski, D. Bojanjac

Faculty of Civil Engineering, University of Zagreb, Croatia, and Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

The Sava River Basin, as the largest river basin in Croatia, is of great importance for the water resources management and hydrological research in Croatia. The general intensification of extreme climatic conditions increases the number of flood events worldwide, which may indirectly affect the morphodynamical behavior in the riverbed. Valuable information on the flood-generation mechanisms may be contained in the flood hydrograph shapes, which can be expressed as continuous functions using a functional data approach [1]. In this research, we performed a preliminary analysis of annual flood events at the gauging station (GS)

Zagreb by applying a clustering mechanism based on functional data to identify a set of representative hydrograph shapes that contain valuable information for future research on scour in the vicinity of bridges.

Flood events that capture river's flood regime can be represented as classical multivariate data or functional data, to be subsequently classified by applying a clustering algorithm. The functional analysis can provide a more objective and reproducible definition of the actual hydrological phenomena than the classical multidimensional analysis because it avoids the subjective selection of a set of hydrograph characteristics. Functional data analysis starts by considering the discrete data observations as part of a finite-dimensional space spanned by a set of basis functions and coefficients that define their linear combination. In addition, the choice of the appropriate order of polynomial segments and the number and placement of the knots are determined to represent the main characteristics of the hydrograph shapes. Finally, the resulting set of coefficients is used as an input for clustering. Since the hydrograph shape clusters can only be validated by their interpretability and usefulness, their evaluation was performed graphically [1,2].

The input data for the analysis was a historical time series of daily discharge data from the Zagreb gauging station in the Sava River basin for the reference period 1960-2019, provided by the Croatian Meteorological and Hydrological Service. The time series consists of 21878 observations (60 years of data). The annual maxima method was applied to the time series to extract independent flood events corresponding to the maximum annual peak discharge values. For each event the flood volume and duration were determined over a fixed window corresponding to the longest duration of the rising and falling limb of all considered flood events, resulting in 16 days before and 38 days after the peak discharge (55 daily observations). The baseflow was separated from the direct flow using the R package "lfstat", with the direct flow component normalized by dividing the ordinate of each hydrograph with the total volume of the flood event. Finally, a functional data analysis was performed by spanning B-spline functions of differing ranks onto the resulting direct flow data observations followed by implementing a k-means clustering algorithm to separate flood events into different clusters. For each cluster, a median was determined to represent three distinct types of hydrograph shapes.

For this analysis the B-spline rank of 25 and the K value of 3 were ultimately chosen. This approach resulted in identifying three distinct types of flood events: (1) slow events described as both elongated rising and falling limbs, (2) intermediate events with steep rising limbs and moderately steep falling limb, and (3) fast events with steep rising and falling limbs of the hydrographs.

This preliminary analysis was conducted in order to exploit the valuable process information contained in flood hydrograph shapes, which can be of great value

for the future research within the R3PEAT project ("Remote Real-time Riprap Protection Erosion Assessment on large rivers", UIP-2019-04-4046) supported by Croatian Science Foundation, that explores influences on the riverbed erosion around structure of bridges crossing large rivers in Croatia [3].

References:

1 Brunner, M. I., Viviroli, D., Furrer, R., Seibert, J., and Favre, A. C. (2018). Identification of flood reactivity regions via the functional clustering of hydrographs. Water Resources Research, 54(3), 1852-1867.

2 Whipple, A. A., Viers, J. H., and Dahlke, H. E. (2017). Flood regime typology for floodplain ecosystem management as applied to the unregulated Cosumnes River of California, United States. Ecohydrology, 10(5), e1817

3 Kovačević, M., Potočki, K., and Gilja, G. (2021, April). The analysis of streamflow variability and flood wave characteristics on the two lowland rivers in Croatia. In EGU General Assembly Conference Abstracts (pp. EGU21-2563).

Automated Methods for Age and Sex Estimation of Adult Panoramic and Individual Dental X-ray Images

D. Milošević, M. Vodanović, I. Galić, M. Subašić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, and School of Dental Medicine, University of Zagreb, Croatia, and University Hospital Centre Split, Croatia

The forensic process tries to answer, among others, the question of identity based on human remains. Forensic odontology is a branch of forensic science focused on the dental system. One of the key steps to determine the identity of a person is to determine the age at the time of death and the sex of the person being processed. The tasks of sex assessment and age estimation can also be used in legal proceedings to protect the rights of people without proper documentation, be it for seeking asylum or when taking care of a found child. Although there are many ways to estimate those properties, teeth and jaw analyses reign supreme with their ratio of indicativeness, non-invasiveness, and durability. Teeth outlast all other tissue when it comes to decomposition, making it a prime forensic target for identification, age, and sex assessment [1]. This feature makes them useful in age estimation for archaeological research, where demographic data and the changes therein can be gathered from important historical sites. X-ray imaging also allows for a non-invasive approach to estimation, avoiding permanent injury to a person or destruction of evidence.

Current estimation methods rely on manual measurements and human estimations [2, 3]. This allows for human error to creep into the results, especially

in disaster situations where the workload for forensic experts far exceeds reasonable amounts. In addition, whereas estimating the age of minors is relatively straightforward due to a wide selection of developmental markers, estimation in adults and seniors still poses a big problem in forensic odontology. Dental corrections, illnesses, and loss of teeth, and the skeletal support structure of the jaw contribute to the problems posed to manual measurement of dental parameters. In such conditions, deep learning can not only automate some tasks but also improve reliability.

Our dataset consists of 4035 anonymized panoramic dental x-ray images. The age range is from 19 to 85 years, with a female to male ratio of 58.7%:41.3%. The samples are collected from multiple locations in Croatia and belong to the collection of the Department of Dental Anthropology School of Dental Medicine University of Zagreb, and they are approved for research purposes by the School's ethics committee. Our samples contain various pathologies and loss of mandibular molars or anomalous molars and teeth, unlike studies usually conducted in the field of forensic odontology. When considered as a dataset of individual teeth, this dataset contains 76416 individual teeth, with 44401 belonging to female and 32015 belonging to male samples. Individual teeth images were extracted from a subset of 2683 manually annotated panoramic dental x-ray images.

The studied models consist of four main parts: the base state-of-the-art convolutional neural network architecture, an additional 1x1 convolutional layer to change the number of feature maps in the final layer of the feature extractor, an optional attention mechanism, and a two-layer fully-connected network with an adjustable number of units in the first fully connected layer. An exhaustive hyperparameter search was used to determine the best performing deep learning model architecture, and depending on the task grid search or random search was used as the search strategy.

As already mentioned, the model consists of four parts. Each part is a proven and well-known component. The goal of this study is to explore the applicability of deep learning methods on the forensic odontology problem of sex assessment. While a custom-made model might achieve marginally better results, those solutions often provide brittle, overfit models.

The first part is the convolutional neural network architecture used as the feature extractor. The following architectures were tested: DenseNet201, InceptionResNetV2, ResNet50, VGG16, VGG19 and Xception. The viability of using ImageNet pretrained network weights was evaluated in preliminary experiments.

Our research has currently developed models for 5 tasks - age estimation from panoramic dental x-ray images, age estimation from individual tooth x-ray images, sex assessment from panoramic dental x-ray images, sex assessment from

individual tooth x-ray images, and tooth type determination.

In the model for age estimation from panoramic dental x-ray images, the mean absolute error achieved is 3.96 years, the median absolute error is 2.95 years and R^2 is 0.8439. Single-tooth estimators, depending on the tooth, achieve mean absolute errors in the range from 6.30 years to 8.68 years and median absolute errors in the range from 4.68 years to 7.33 years, with R^2 ranging from 0.3277 (down-3) to 0.5541 (down-7).

For sex assessment, our model achieves an accuracy of 96.87% ± 0.96% on our dataset. The performance decreases with sample age, so while it achieves 100% accuracy on the age group of 20 to 30 years old, that accuracy declines to 95% for the age bracket of 60 to 70 years old. In contrast, sex assessment of individual teeth achieves an accuracy of just 72.68%. Both the general and specialized-by-tooth-type models perform similarly. However, when evaluated on a subset of verified healthy and unaltered teeth, the accuracy rises to 85%, indicating that the inclusion alterations and diseases impede model performance.

Acknowledgment: This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References:

1 J. C. Dudar, S. Pfeiffer, and S. Saunders, "Evaluation of morphological andhistological adult skeletal age-at-death estimation techniques using ribs," Journal of Forensic Science, vol. 38, no. 3, pp. 677–685, 1993.

2 S. I. Kvaal, K. M. Kolltveit, I. O. Thomsen, and T. Solheim, "Age estimation of adults from dental radiographs," Forensic Science International, vol. 74, pp. 175–185, July 1995.

3 R. Cameriere, L. Ferrante, and M. Cingolani, "Variations in Pulp/Tooth Area Ratio as an Indicator of Age: a Preliminary Study," Journal of Forensic Science, vol. 49, pp. 1–3, Feb. 2004. Publisher: ASTM International.

Starguider camera - Measuring telescope pointing

T. Šarić

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture

Cherenkov Telescope Array (CTA) is a next-generation ground-based observatory for very high-energy gamma-ray astronomy. It consists of arrays of Imaging Atmospheric Cherenkov Telescopes aiming to improve sensitivity level by an order of magnitude at 1 TeV compared to current instruments, improve angular resolution, the field of view, and provide energy coverage for photons from 20

GeV to at least 300 TeV. Three different sized telescopes (LST, MST, SST) are sensitive in different energy regimes to meet the necessity of a wide energy range which maximizes the chance of serendipitous detection.[1]

Starguider(SG) camera is a pointing device for precise and independent measurements of the LST-1 telescope pointing. It is located at the center of the dish of the telescope where it sees part of the night sky and the part of the Cherenkov camera. In order to make use of SG camera images, one needs to develop software that can analyze its images. Regarding the part with the night sky, software must recognize what part of the sky SG sees, which means to detect sources as potential stars, match them to catalog stars and map every pixel of an image to sky coordinates. We report an average obtained solving accuracy of 1.6 arcseconds measured as the offsets of a star in the image from the catalog position.

The other part of the SG image consists of a Cherenkov camera which among others contains 6 LEDs. Software needs to detect the centroid of each LED, fit a circle to LEDs, and eventually map the center of the circle i.e. the center of the field of the LST to a sky coordinates obtained in the previous step. Finally, one can compare the nominal(planned) pointing of the telescope with the pointing measured by Starguider to calculate telescope misspointing.

Software for measuring LST pointing by analyzing SG images is dockerized, deployed on a server, and switched on for real-time data analysis. It analyses every new image taken, calculates various useful statistics, plots graphs, and posts them on an LST webpage where shifters can monitor them and promptly intervene in case of any deviation of telescope pointing. Results are also saved in the database for an offline, more detailed analysis. We report an average measured misspoiting of 4.259 arcminutes with an std of 1.549. The result is rather large, we assume due to insufficiently precise SG displacement correction (SG is shifted from the center of the mirror carrier in x and y direction as well as rotated around the y-axis which causes the effect of parallax). The second contribution to an error comes from inaccurate detection of the centroid of LEDs part from the oversaturated pixels and part from the reflection of light on the targets for distance meters which causes overflow of pixel intensity between LEDs and targets. For future work, we are exploring another approach to measure telescope pointing by determining the fixed angle between the optical axis of the telescope and the SG camera. This approach, if successful, could overcome problems with LEDs as well as with mentioned parallax effect.

References:

1 https://arxiv.org/abs/1709.07997

Author Index

Notes