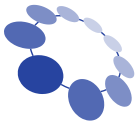
Third International Workshop on Data Science

Abstract Book

IWDS 2018



RESEARCH UNIT FOR DATA SCIENCE

Centre of Research Excellence for Data Science and Cooperative Systems

DATACROSS project is funded by the European Union through the European Regional Development Fund

Europska unija
Zajedno do fondova EU

Ministarstvo znanosti i obrazovanja

Zagreb, October 16, 2018

Abstract Book

Third International Workshop on Data Science

Zagreb, Croatia, October 16, 2018

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
IEEE Croatia Section

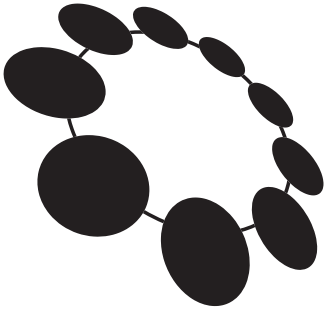
Editors

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

Welcome Message

On behalf of the Organizing Committee it is with great pleasure that we extend a warm welcome to all participants of the Third International Workshop on Data Science (IWDS 2018) in Zagreb. The workshop is organized by the Research Unit for Data Science (RUDS), Centre of Research Excellence for Data Science and Cooperative Systems. Faculty of Electrical Engineering and Computing, University of Zagreb is the host organization of the IWDS 2018.

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 2018 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 2018.

Zagreb, October 2018

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 Committee

Enis Afgan, Croatia
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 Pripuđić, 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

Finance Chair	Vedran Podobnik, Croatia
Registration Chair	Zvonko Kostanjčar, Croatia
Publications Chair	Ivana Zadro, Croatia
Industry Liaison	Mihaela Vranić, Croatia
Local Arrangements Chair	Damir Pintar, Croatia
Secretariat	Ana Kroflin, Croatia

	Tuesday, 16th October, 2018, Grey Hall, FER Conference Center
08:30 – 09:00	Registration of Participants
09:00 – 09:05	**Opening** *Nikola Mišković, Vice-Dean for Research, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*
09:05 – 09:10	**Introduction** *Sven Lončarić, University of Zagreb, Croatia and* *Tomislav Šmuc, Ruđer Bošković Institute, Croatia*
09:10 – 09:50	Plenary Lecture **Mining Evolving Social Networks** *João Gama, Laboratory of Artificial Intelligence and Decision Support, and Faculty of Economics, University of Porto, Portugal*
09:50 – 10:15	Coffee Break
10:15 – 12:00	**An Overview of Research Activities in Data Science**
	Research Session 1 **Multimedia Data Processing and Data Management** *Co-Chairs: Jan Šnajder, University of Zagreb, Croatia and* *Ivica Kopriva, Ruđer Bošković Institute, Croatia*
	Research Session 2 **Machine Learning and Data Mining** *Co-Chairs: Tomislav Šmuc, Ruđer Bošković Institute, Croatia and* *Bojana Dalbelo-Bašić, University of Zagreb, Croatia*
	Research Session 3 **Heterogeneous Computing and Advanced Cloud Services** *Co-Chairs: Karolj Skala, Ruđer Bošković Institute, Croatia and* *Zlatan Car, University of Rijeka, Croatia*
	Research Session 4 **Multidisciplinary Data Intensive Applications** *Co-Chairs: Vedran Podobnik, Kristian Vlahoviček, University of Zagreb, Croatia and* *Nikola Godinović, University of Split, Croatia*
12:00 – 12:30	**Poster Session** (Posters will be on display from 9:00.)
12:30 – 13:30	Lunch Break (FER Gallery)

	Tuesday, 16th October, 2018, White Hall, FER Conference Center : Internal meetings
13:30 – 14:30	**Industry Advisory Board meeting**
15:00 – 16:00	**Scientific Advisory Board meeting**

Spline-like Chebyshev Polynomial Representation for Compressed Sensing

T. Vlašić, J. Ivanković, D. Seršić, A. Tafro

Faculty of Electrical Engineering and Computing, University of Zagreb

Signal acquisition and reconstruction is the heart of signal processing and sampling theorems provide the bridge between the continuous and the discrete-time worlds. The most widely used sampling theorem is often attributed to Shannon and gives a sufficient condition, namely bandlimitedness, for exact sampling and interpolation formula. The sampling rate, at twice the maximum frequency present in the signal, is called the Nyquist rate. Using the same sampling concept, non-bandlimited signals will be projected onto the bandlimited subspace. Vetterli et al. (2002) introduced the term "rate of innovation", which is the degrees of freedom of signal per unit of time. Examples of signals with a finite rate of innovation (FRI) are nonuniform splines and piecewise polynomials, which are non-bandlimited parametric signals. One can perfectly reconstruct those signals using sampling at the rate of innovation and the algorithms that are more complex than the standard one [1].

Compressive sampling or compressed sensing (CS) [2] is a technique for signal sampling below the Nyquist rate, based on the assumption that the signal is sparse in some transform domain (e.g. DCT, DFT, DWT). If a signal has a sparse representation s in some transform domain Ψ, one can discard the small coefficients without much perceptual loss. Therefore, s is the vector of coefficients with all but the largest S coefficients set to zero. In CS, coefficients are obtained from optimization operations on measurements y that are linear projections of the signal, without the acquisition of the whole signal x. Ideally, measurement matrix Φ is designed to reduce number of measurements as much as possible, while allowing the recovery of a wide class of signals. While CS, on one hand, reduces the number of measurements and the sampling rate, on the other hand, it increases the computational complexity of the signal recovery. CS reconstruction processes can be observed as linear inverse problems that often occur in numerous image processing tasks. CS acquires a signal in a compressed form that is appropriate for storage, transmission and further processing.

Signals with finite rate of innovation can be seen as "sparse" in the time domain. However, differently from the compressed sensing framework, this domain is not discrete: the innovation time may assume arbitrary real values. Yet, assuming that these innovations fall on some discrete grid (e.g. uniformly divided interval) know a priori, one may address FRI interpolation problem as an optimization problem proposed in CS [3]. It was shown that new paradigm provides the parameters of the innovation with "overwhelming" probability using even less

measurements. CS changed an outlook on the non-bandlimited parametric FRI signal acquisition and reconstruction.

Many classes of signals and images are compressible in bases like DCT, DWT, etc. Therefore, these bases are often used in the CS paradigm as a transform domain. Frequently, signals and images have to be analyzed and processed (filtering, edge detection, interpolation, segmentation, etc.). Digital signal processing is performed on discrete data sets (samples) which have to be decoded from a compressed form. Decoded images often consists of blocking artifacts (JPEG) which present an issue in image processing. The question is: can we acquire signals and images using the CS theory and conduct signal processing on obtained coefficients (parameters) rather than decoded samples? Is it possible to avoid the blocking artifacts within the CS recovery process?

We propose use of the Chebyshev polynomials on intervals for an efficient representation of one-dimensional continuous signals and images. The DCT is suitable as a transform domain in the CS theory because of its strong energy compaction property. Since the Chebyshev polynomial coefficients are the DCT of the function values on the nonuniform grid, we can assume that they approximate a wide class of signals in a similar manner. Since the Chebyshev polynomials are parametric signals, the signal processing can be conduct directly on its parameters. We propose the parametric model that fits into the CS setup and offer a new paradigm for efficient processing of the analog data on a digital computer.

We propose signals or images to be divided into equal-size intervals or blocks and piece-wise approximated by the Chebyshev polynomials. If every interval is approximated independently, it is likely to expect discontinuities on the edges of the intervals. Splines, which are smooth, well behaved and continuous at every point of piece-wise defined polynomials, are frequently used to represent continuous-time signals. If P is the highest polynomial degree, the spline approach asserts that polynomial segments are patched together so that the interpolating function and all its derivatives up to order $P-1$ are continuous at all joining intersections. Subsequently, values and $P-1$ derivatives are equalized at the edges of the intervals. This leads to a system of equations that has an exact solution and consequently does not fit into the CS paradigm. Therefore, a spline-like interpolation model is proposed. We equalize values of only a few derivatives of each pair of neighboring approximation functions. It results in a system of equations that is under-determined providing an overall framework for finding sparse solutions by optimization. Consider the fact that the proposed method still brings a desired level of continuity at the intersections of intervals or blocks.

Due to simplicity of the Chebyshev polynomials, there exist closed form solutions for its derivatives. The given conditions on the polynomial functions are

embedded in constraint matrix. The constraint matrix is introduced to the linear optimization problem when recovering the transform of the input signal. Among infinitely many solutions that correspond to the actual measurements, we choose those representations that fall into the null-space or kernel of constraint matrix. We propose an acquisition system with channels consisting of demodulation, integration and uniform sampling. Each channel corresponds to one measurement, i.e. one row of the measurement matrix. Input signal $x(t)$ is modulated by uniformly distributed pseudo-random signal $r(t)$. In our example, the pseudo-random signal is piece-wise constant and uniformly divided into subintervals. The same pattern of values is periodically repeated for each subsequent interval. The modulated signal passes through an ideal integrator that serves as a low-pass filter in the frequency domain. At the beginning of every interval, the integrator is reset to zero. Finally, a measurement sequence y is obtained by sampling the integrator's output at the end of each interval at a sampling rate lower than the Nyquist rate of the input signal. The system consists of channels that differ only in pseudo-random signal $r(t)$. Sampled data y is utilized in the optimization problem to reconstruct the input signal $x(t)$.

The last step of the CS process is the reconstruction of the input signal from the sampling sequence through a suitable signal reconstruction algorithm. The recovery of the input signal from measurements y is an ill-posed problem in general; however, the compressibility of the signal makes the recovery feasible. The recovery of the set of Chebyshev polynomial coefficients can be achieved through the L1 optimization problem. The optimization problem for the proposed system can be solved with traditional convex programming techniques or iterative greedy algorithms like Orthogonal Matching Pursuit. The simulated experiments have proven feasibility of the proposed concept. Analog signals, which were not band limited, were successfully restored from the reduced set of measurements in the presence of additive noise. Resulting parametric models of the analog input can be directly processed on a digital computer, without a need of conversion to samples. Results show that the proposed paradigm is feasible and has promising application in measurement, representation and direct processing of analog signals.

References:

[1] M. Vetterli, P. Marziliano and T. Blu, "Sampling signals with finite rate of innovation," IEEE Trans. Sig. Proc., vol. 50, pp. 1417-1428, June 2002.

[2] D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory, vol. 52, pp. 1289-1306, Apr. 2006.

[3] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano and L. Coulot, "Sparse Sampling of Signal Innovations," IEEE Sig. Proc. Magazine, vol. 25, no. 2, pp. 31-40, Mar. 2008.

Notes

Project co-leaders:

Prof. dr. sc. Sven Lončarić
sven.loncaric@fer.hr

Prof. dr. sc. Ivan Petrović
ivan.petrovic@fer.hr

DATACROSS Project Office:

Mirjana Stjepanović
mirjana.stjepanovic@fer.hr

Ana Kroflin
ana.kroflin@fer.hr

Address:

University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3, 10 000 Zagreb
01/6129-542

https://accross-datascience.zci.hr/datacross
www.strukturnifondovi.hr