Journal of Machine Learning Research 16 (2015) 1677-1681 Published 9/15

Preface to this Special Issue

Alex Gammerman ALEX@CS.RHUL.AC.UK
Vladimir Vovk V.VOVK@RHUL.AC.UK
Computer Learning Research Centre, Department of Computer Science
Royal Holloway, University of London

This issue of JMLR is devoted to the memory of Alexey Chervonenkis. Over the period of a dozen years between 1962 and 1973 he and Vladimir Vapnik created a new discipline of statistical learning theory—the foundation on which all our modern understanding of pattern recognition is based. Alexey was 28 years old when they made their most famous and original discovery, the uniform law of large numbers. In that short period Vapnik and Chervonenkis also introduced the main concepts of statistical learning theory, such as VC-dimension, capacity control, and the Structural Risk Minimization principle, and designed two powerful pattern recognition methods, Generalised Portrait and Optimal Separating Hyperplane, later transformed by Vladimir Vapnik into Support Vector Machine—arguably one of the best tools for pattern recognition and regression estimation. Thereafter Alexey continued to publish original and important contributions to learning theory. He was also active in research in several applied fields, including geology, bioinformatics, medicine, and advertising.

Alexey tragically died in September 2014 after getting lost during a hike in the Elk Island park on the outskirts of Moscow. Vladimir Vapnik suggested to prepare an issue of JMLR to be published at the first anniversary of the death of his long-term collaborator and close friend. Vladimir and the editors contacted a few dozen leading researchers in the fields of machine learning related to Alexey's research interests and had many enthusiastic replies. In the end eleven papers were accepted. This issue also contains a first attempt at a complete bibliography of Alexey Chervonenkis's publications.

Simultaneously with this special issue will appear Alexey's Festschrift (Vovk et al., 2015), to which the reader is referred for information about Alexey's research, life, and death. The Festschrift is based in part on a symposium held in Pathos, Cyprus, in 2013 to celebrate Alexey's 75th anniversary. Apart from research contributions, it contains Alexey's reminiscences about his early work on statistical learning with Vladimir Vapnik, a reprint of their seminal 1971 paper, a historical chapter by R. M. Dudley, reminiscences of Alexey's and Vladimir's close colleague Vasily Novoseltsev, and three reviews of various measures of complexity used in machine learning ("Measures of Complexity" is both the name of the symposium and the title of the book). Among Alexey's contributions to machine learning (mostly joint with Vladimir Vapnik) discussed in the book are:

- derivation of necessary and sufficient conditions for the uniform convergence of the frequencies of events to their probabilities (later developed into necessary and sufficient conditions for the uniform convergence of means to expectations);
- introduction of a new characteristic of classes of sets, which they called capacity (емкость) and which was later renamed as VC-dimension;

- development of powerful pattern recognition algorithms, Generalized Portrait and Optimal Separating Hyperplane;
- applying the theory of machine learning to diverse fields; e.g., a computer system using methods of machine learning was developed and installed at the world's largest open gold pit in Murun-Tau (Uzbekistan), which won him the State Prize of the USSR.

This Special Issue opens with the paper by Vladimir Vapnik and Rauf Izmailov "V-matrix method of solving statistical inference problems", which proposes new ways of solving learning problems such as estimating conditional probabilities. The authors ask whether the new methods can replace SVM in the problem of pattern recognition. Solving the problem of pattern recognition via estimating conditional probabilities might appear to contradict Vapnik's [1995, 1998] Imperative that the problem of interest (in this case pattern recognition) should be solved directly, without solving a more general problem (in this case estimating conditional probabilities) as an intermediate step. However, the authors explain that there is no real contradiction.

The paper "Batch learning from logged bandit feedback through Counterfactual Risk Minimization" by Adith Swaminathan and Thorsten Joachims extends Vapnik and Chervonenkis's Structural Risk Minimization principle to the situation where only partial feedback, determined by the prediction, is available. Such situations are ubiquitous in, e.g., advertisement placement, an active area of Alexey's research during the last years of his life.

The next paper, "Optimal estimation of low rank density matrices" by Vladimir Koltchinskii and Dong Xia, concerns Alexey's other major interest, quantum mechanics, which the authors mention at the beginning of their contribution. This interest is not reflected in Alexey's bibliography (published at the end of this Special Issue), and we know about it from reminiscences of his colleagues and relatives and from his technical report (Chervonenkis, 2001), in which he computes the covariance function for the solution to Schrödinger's equation. The paper by Koltchinskii and Xia is devoted to the estimation of density matrices, describing states of quantum systems, which has important applications in quantum tomography.

The paper "Fast rates in statistical and online learning" by van Erven, Peter D. Grünwald, Nishant A. Mehta, Mark D. Reid, and Robert C. Williamson explores conditions that make fast learning possible finding unexpected similarities between two styles of learning, statistical and online. It is interesting that Vapnik and Chervonenkis started their joint work in a non-statistical setting (Chervonenkis, 2015), although they quickly moved to their well-known statistical one, which is now standard in machine learning. The authors show us that the difference between the two styles of learning is smaller than it appears.

In their paper "On the asymptotic normality of an estimate of a regression functional" László Györfi and Harro Walk deal with the standard problem of regression in statistical learning theory but are interested in the quality (as measured by the square loss function) of the regression function rather than the regression function itself. They prove a surprisingly robust result about the asymptotic distribution of the main component of this measure of quality.

Pierre Bellec and Alexandre B. Tsybakov in "Sharp oracle bounds for monotone and convex regression through aggregation" consider important restricted versions of regression, in which the regression function is assumed to be either monotone (isotonic regression)

or convex. In evaluating their procedures the authors follow Vapnik and Chervonenkis's criterion of minimax loss, their main tools are different kinds of predictor aggregation, and their non-asymptotic performance guarantees are sharp.

The paper "Exceptional rotations of random graphs: a VC theory" by Louigi Addario-Berry, Shankar Bhamidi, Sébastien Bubeck, Luc Devroye, Gábor Lugosi, and Roberto Imbuzeiro Oliveira develops a fascinating analogue of the Vapnik–Chervonenkis statistical learning theory adapting it to random graphs.

The next paper, "Semi-supervised interpolation in an anticausal learning scenario" by Dominik Janzing and Bernhard Schölkopf, represents the area of causal inference and sheds new light on the situations in which seeing unlabelled observations does not provide any useful information (and so semi-supervised learning does not work). As the second author modestly says in a slightly different context elsewhere (Schölkopf, 2014), such results may not be as beautiful as those in the field that Alexey co-founded; however, this is compensated by their practical and philosophical importance.

It appears that unsupervised learning was one of the few fields of machine learning in which Alexey did not work directly, despite its importance in many applications: no one mind, even as versatile as his, can embrace everything. In their "Towards an axiomatic approach to hierarchical clustering of measures", Philipp Thomann, Ingo Steinwart, and Nico Schmid study the foundations of unsupervised learning. They show how the user's choice of a "clustering base" in conjunction with several natural axioms determines a clustering method.

Mark Herbster, Stephen Pasteris, and Massimiliano Pontil's "Predicting a switching sequence of graph labellings" is another paper devoted to online learning. The authors design new online prediction algorithms on graphs that can cope with switching labellings and multitask prediction problems.

The last research paper in this Special Issue is Vladimir Vapnik and Rauf Izmailov's "Learning using privileged information: Similarity control and knowledge transfer", which complements the standard protocol of statistical learning with an Intelligent Teacher providing the Student with privileged information. Such information is present in many real-world applications of machine learning and can be very useful.

The last part of the Special Issue is Alexey Chervonenkis's bibliography. His publications are listed in the chronological order, starting from the fundamental papers by Vapnik and Chervonenkis on the method of Generalized Portrait and the foundations of statistical learning theory, and then branching into countless fields including applied linguistics, geology, medicine and bioinformatics, and advertisement placement. They attest to his great role as discoverer and inventor. His tragic death a year ago was a great loss not only to his relatives, friends, and colleagues, who remember his wonderful warmth as a person, but also to the whole machine learning community and science in general.

References

Alexey Chervonenkis. Covariance function for Shrodinger equations decision. Technical Report CLRC-TR-01-03, Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, UK, April 2001. URL `www.clrc.rhul.ac.uk/publications/files/tr0103.ps`. Accessed in September 2015.

Alexey Chervonenkis. Chervonenkis's recollections. In Vovk et al. (2015), pages 3–8.

Bernhard Schölkopf, 2014. URL `http://people.tuebingen.mpg.de/bs/chervonenkis.html`. Post in memory of Alexey Chervonenkis. Accessed in September 2015.

Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995. Second edition: 2000.

Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

Vladimir Vovk, Harris Papadopoulos, and Alex Gammerman, editors. *Measures of Complexity: Festschrift for Alexey Chervonenkis*. Springer, Berlin, 2015.

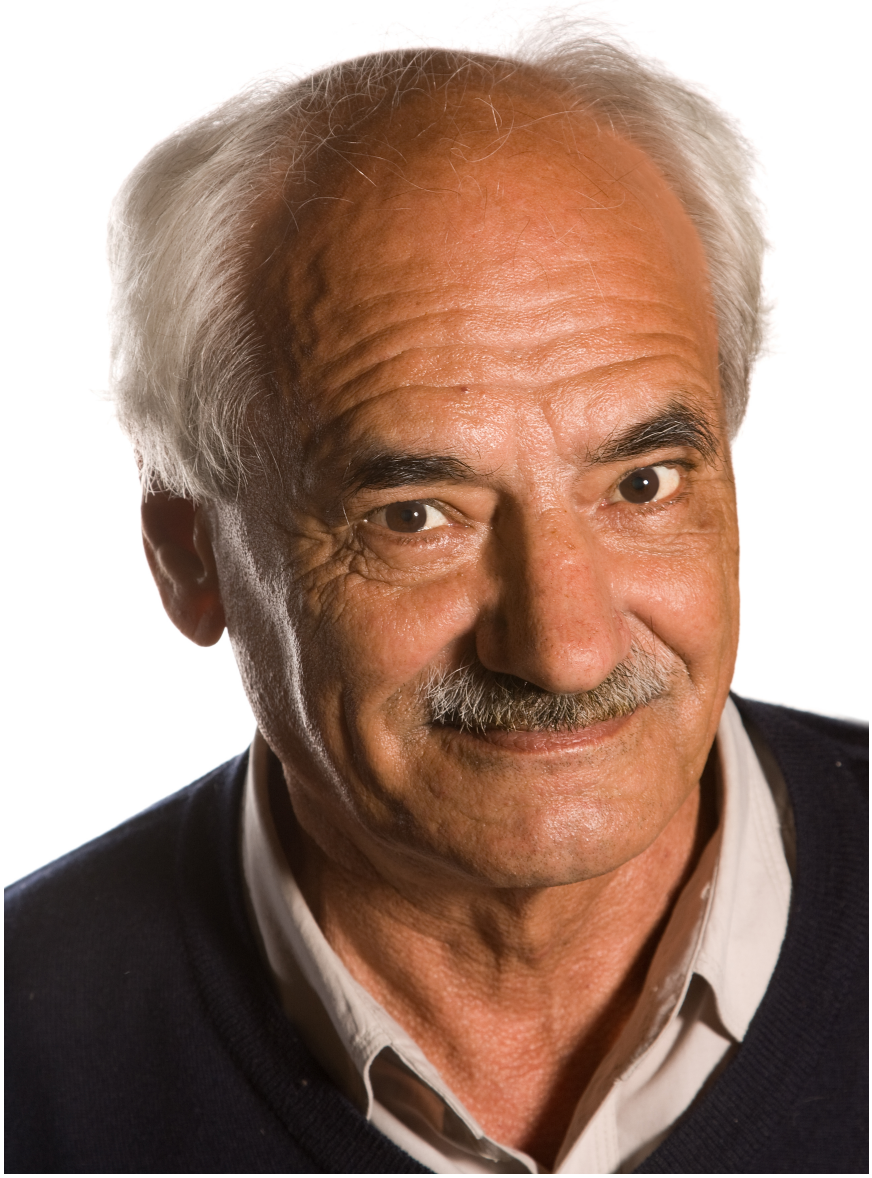

Figure 1: Alexey Chervonenkis (1938–2014)