Vladimir Vovk · Harris Papadopoulos
Alexander Gammerman *Editors*

Measures of Complexity

Festschrift for Alexey Chervonenkis

Springer

Measures of Complexity

Vladimir Vovk · Harris Papadopoulos
Alexander Gammerman
Editors

Measures of Complexity

Festschrift for Alexey Chervonenkis

Springer

Editors
Vladimir Vovk
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey
UK

Alexander Gammerman
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey
UK

Harris Papadopoulos
Department of Computer Science and Engineering
Frederick University
Nicosia
Cyprus

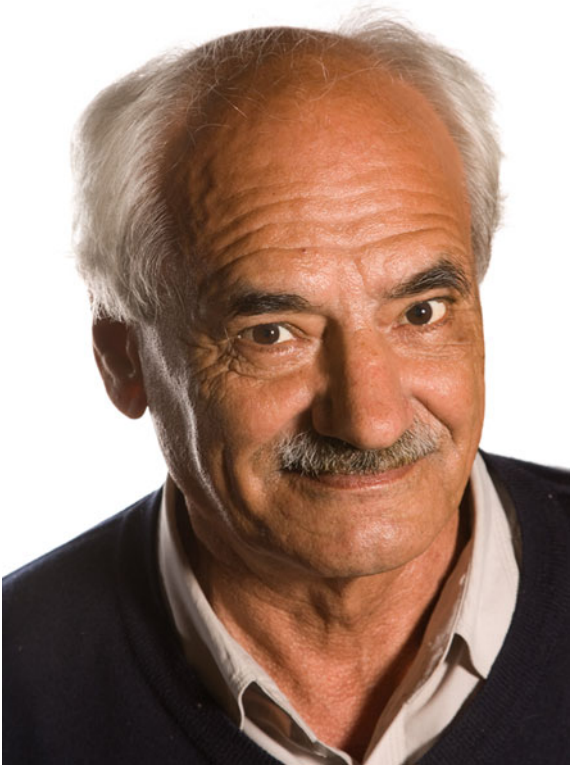
ISBN 978-3-319-21851-9 ISBN 978-3-319-21852-6 (eBook)
DOI 10.1007/978-3-319-21852-6

Library of Congress Control Number: 2015946591

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)



Alexey Chervonenkis (1938–2014)

Preface

In our media-centered age obsessed with various semi-known and unknown personalities and celebrities, the life and work of one of the founders of modern machine learning, Alexey Chervonenkis, somehow remains largely unknown. Alexey celebrated his 75th anniversary in 2013, and several of his colleagues organized a symposium devoted to his life and work. The symposium was held in Paphos, Cyprus, on October 2, 2013, and was called "Measures of Complexity." To some degree, the present volume is an outcome of that meeting; some of the chapters (such as Chap. 13 by Alexey Chervonenkis and Chaps. 4 and 14 by Richard Dudley) are based on the talks delivered by their authors at the symposium. But the vast majority of the chapters were prepared specifically for this volume.

Two years earlier the machine learning community had celebrated the 75th anniversary of Alexey's close friend and co-author Vladimir Vapnik, and the Vapnik Festschrift was published recently as [1]. Compared to the Vapnik Festschrift, this volume is somewhat less theoretical. It contains four parts: history; reviews of different notions of complexity; discussion of possible refinements of VC bounds; and technical contributions (in fact quite a few of them are reviews of specialized areas of machine learning, or contain such reviews). The main strength of this volume might be not so much in its original results (although there are a few impressive new results in Part IV) as in being a source of motivation and information for Ph.D. students and new researchers entering the field.

Egham — Vladimir Vovk
Nicosia — Harris Papadopoulos
Egham — Alex Gammerman
June 2015

Reference

1. Schölkopf, B., Luo, Z., Vovk, V. (eds.): Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik. Springer, Berlin (2013)

Short Biography of Alexey Chervonenkis

Alexey Chervonenkis made a long and outstanding contribution to the area of pattern recognition and computational learning. His book [9] on pattern recognition co-authored with Vladimir Vapnik was published in 1974 and made him an established authority in the field. His most important specific contributions (most of them also joint with Vladimir Vapnik) to statistical learning theory include:

- Derivation of necessary and sufficient conditions for the uniform convergence of the frequency of an event to its probability over a class of events (a result that was later developed into necessary and sufficient conditions for the uniform convergence of means to expectations).
- The introduction of a new characteristic of a class of sets, later called the VC dimension.
- Development of a pattern recognition algorithm called "generalized portrait," which was later further developed into the well-known Support Vector Machine.
- Development of principles and algorithms for choosing the optimal parameters depending on the amount of available empirical data and the complexity of the class of decision rules for the problems of pattern recognition and regression. A computer system using these methods was developed and installed at the world's largest open gold pit in Murun-Tau (Uzbekistan).

Some of his theoretical results served as the foundation for many practical machine learning algorithms. Here is a short biography of Alexey Chervonenkis.

Early Years

Alexey Yakovlevich Chervonenkis (Алексей Яковлевич Червоненкис in Cyrillic) was born in 1938 in Moscow, USSR. He spent all his childhood in Moscow, apart from two years during World War II (October 1941–December 1943), when his family was evacuated to and lived in Omsk in Siberia. His father was an electrical engineering scientist, and both his parents worked for the National Energy System

of the USSR. His grandfather M.R. Chervonenkis was well known in Russia as one of the 12 Jews who were elected deputies (MPs) of the very first State Duma (Parliament) of Russia in 1906.

In 1946 Alexey went to school, and after graduating in 1955 he became a student of the Moscow Institute of Physics and Technology (MIPT)—a state university, informally called Phystech and sometimes referred to as the "Russian MIT." It was one of the best higher education establishments in the USSR (and still is one of the best in the former USSR). In addition to the usual lecture courses, the students also spent their time in various industrial research laboratories, and Alexey was attached to the Institute of Automation and Remote Control of the Soviet Academy of Sciences, ICS (the Institute was later renamed as the Institute of Control Sciences).

Alexey's final year project at the MIPT (and ICS) was devoted to the construction of a digital-to-analog converter. One of the problems he faced in this project was how to reconstruct a continuous function from discrete measurements with a predetermined accuracy.

Institute of Control Sciences

After graduation from the MIPT in 1961 Alexey stayed to work at the same Institute (ICS) but at a different laboratory, and was employed there for the rest of his life. At that time the ICS had already established itself as an intellectual powerhouse, in large degree due to Vadim Trapeznikov, who was appointed the Institute's Director in 1951. Among Trapeznikov's tools were fostering rivalry within and between laboratories and fighting anti-semitism in the 1950s and 1960s.

Alexey's first year or so at the ICS was devoted to designing a light organ. The Institute was contracted to make this device for a show at the Soviet Industrial Exhibition in London in 1961. The work started at the beginning of 1961 and after a few months of very intensive work Alexey found a way to control the source of light—a film projector with a xenon lamp. He made a feedback implementation using a photocell as a sensor of the light intensity. The device was exhibited in London and later at the Exhibition of Achievements of the National Economy in Moscow.

During his university years Alexey became interested in problems of cybernetics. The topic was first outlined by Norbert Wiener in his 1948 book [12]. But the real boom in this field started in 1958, when the book was translated into Russian. The subject also had a stigma of being persecuted by Communist Party apparatchiks since its roots were not in Marxism and, moreover, its philosophical foundations were not clear. But in the 1960s under Khrushchev it became incredibly popular and was considered to be a tool to solve all technical, industrial, and military, and also some other problems in society. Mark Aizerman, who was working at the Institute of Automation and Remote Control and also lecturing at the MIPT, gave a series of lectures on this subject when Alexey was a student. So, it is not surprising that when Alexey finished his work with the light organ, he decided

to concentrate on problems of cybernetics and, in particular, problems of pattern recognition; he was given an opportunity to do so as a Ph.D. student in the laboratory of Aleksandr Lerner of the same Institute. Another Ph.D. student, Vladimir Vapnik from Tashkent, also joined the same laboratory. Vladimir graduated from a university in Uzbekistan and started working in one of the research institutes in Tashkent. Aleksandr Lerner, while on his business trip to Tashkent, was persuaded to take the promising young researcher to Moscow for postgraduate study. The idea was that after receiving his Ph.D. the student would go back to Tashkent and enhance the local research community. One way or another Alexey and Vladimir started their joint work on problems of pattern recognition.

Generalized Portrait

In 1962–1964 they invented a new method of prediction and called it "Generalized Portrait." The algorithm constructed a hyperplane to separate classes of patterns. The method of Generalized Portrait could be reduced to work with scalar products of input vectors. At the time the Institute had no digital computers, only analog ones. This created a problem with inputting the initial data. They did this by calculating the scalar products by hand (or using calculators) and inputting them into the analog computers by adjusting corresponding resistors. Later, starting from 1964, the Institute acquired digital computers, and the method of Generalized Portrait was implemented to solve many different recognition problems in geology, meteorology, and other fields.

Uniform Convergence and VC Dimension

In connection with the Generalized Portrait algorithm, Alexey and Vladimir faced the following problem. If there is a decision rule that makes no errors on the training set, why is there a hope that the percentage of errors on new data will also be zero or small? V. Vapnik, L. Dronfort, and A. Chervonenkis introduced the notion of learning algorithms with *full memory*—that is, algorithms that make no errors on the training set. It turned out that for a finite number N of possible decision rules, the probability of making an error on new data is, with high probability, less than or approximately equal to the ratio of $\log N$ to the length of the training sequence l. For points in n-dimensional space with coordinates allowed to take only a finite number k of values and for linear decision rules, the number N of different rules may be bounded by a polynomial of k with degree equal to the squared dimension n^2, and the bound on the probability of error becomes proportional to the ratio of n^2 to l. (For further details, see the first pages of Alexey's Chap. 1 in this volume.) This result was published in 1964 [6, 11], but the paper [11], probably

never translated into English, was prepared for a conference of young specialists which took place in the spring of 1963; the authors' names were listed, as usual, in alphabetical order (according to the Russian alphabet). The proof used the idea, which was much later reinvented by other authors, of the PAC—probably approximately correct—setting in statistical learning theory.

A key idea came from Mikhail Bongard, who may have been the first to state the problem of learning as the problem of choosing a decision rule out of a set of possible decision rules. In any case, Vapnik and Chervonenkis learned this idea from him.

The Vapnik–Dronfort–Chervonenkis result was criticized by other researchers as covering only a very special case: it dealt with a finite number of rules, and no errors on the training set were allowed. This criticism led its authors to consider the notion of the growth function, which is applicable to infinite sets of decision rules. If the growth function grows as polynomial, then one can get acceptable bounds on the probability of error. Otherwise no nontrivial distribution-independent bounds can be found. This result was obtained in 1966 and first published in 1968 [7]. Further they found that only two cases are possible—polynomial or exponential growth—and in the former case the degree of the polynomial is determined by the largest sample size l at which the growth function is still 2^l. This number was later called the *VC dimension* of the class of decision rules.

The next step was to generalize the result to the case where errors on the training set are allowed. This led to the much more general problem of uniform convergence of frequencies to probabilities over a class of events, known as the generalized Glivenko–Cantelli problem. Inspired by the notion of the growth function, Alexey and Vladimir obtained necessary and sufficient conditions for such convergence. Along the way, they also obtained bounds on the probability of error in the general case. In the case of necessary and sufficient conditions, however, the growth function had to be replaced by the entropy of the class, where the maximum was replaced by the log expectation (the expected value of the logarithm). These results were obtained in 1966–1967, were published in the Proceedings of the USSR Academy of Sciences in 1968 [7], and their proofs were published in 1971 [8].

In 1971 Alexey passed a viva and was awarded a Ph.D. degree. His thesis was devoted to the problem of uniform convergence and its applications to machine learning.

In 1974 Alexey and Vladimir published the book "Theory of Pattern Recognition" [9] containing all the results described above, their improved versions, and much more. Later, in the 1970s and 1980s, these results were developed in several directions. In particular, they discovered (1981, [10]) necessary and sufficient conditions for the uniform convergence of means to expectations over a class of functions. They also proposed to use bounds on the probability of error to choose the optimal complexity of a decision rule. This approach is known as the *principle of structural risk minimization*. Algorithms based on these developments were published in the book "Algorithms and Programs for Dependency Reconstruction" (1984, [4]).

Much later, after emigrating to the USA, Vladimir Vapnik returned to the idea of representing the data by scalar products between vectors, similarly to what he and Alexey did when implementing Generalized Portrait on analog computers. He combined this idea with using kernels as a generalization of scalar products, and this led to the method of Support Vector Machines—a very successful algorithm for pattern recognition and regression estimation.

Murun-Tau: Ore Deposit Exploration

In 1968, Alexey started his collaboration with the Institute of Geology of Ore Deposits, Petrography, Mineralogy, and Geochemistry in Moscow. Ten years earlier the giant Murun-Tau gold deposit had been discovered in the Kyzyl Kum Desert in Uzbekistan. It was estimated that the deposit contained about 170 million ounces of gold, and the Institute started development of an automatic system for contouring ore deposits with a particular application to the Murun-Tau deposit.

Alexey with his colleagues from the Institute considerably improved the accuracy of estimates of gold concentration. To estimate the concentration of gold, initially a large number of 10-meter-deep boreholes arranged in a 5×5 square lattice were drilled in a quarry. The samples extracted from the boreholes were used to estimate the average concentration of gold in the "zone of influence" of each borehole (its Voronoi cell, a $5 \times 5 \times 10$ rectangular box), and the zone of influence was regarded as ore and processed only if the estimate exceeded some threshold; otherwise, the zone of influence was regarded as waste. The traditional estimation method was trivial and simply took the measurement at the borehole as the estimate. But the accuracy of measurements was quite low and, besides, the concentration of gold in a narrow borehole could be very different from the average concentration in the zone of influence; therefore, to make the estimates more accurate one had to take into account the measurements in the nearby boreholes, and their effect depended on the correlations between the concentrations at different points. Estimating the three-dimensional correlation function was the most difficult part of the problem. Alexey successfully overcame this difficulty, and a computer system was developed and has been used in Murun-Tau ever since. The results of this work were very impressive, and in 1987 Alexey and his colleagues were awarded one of the highest prizes in the USSR: the State Prize.

University of London

Alexey Chervonenkis's first visit to the University of London was in Autumn 1998. He came to celebrate the establishment of a new research centre, CLRC (Computer Learning Research Centre), at Royal Holloway, University of London, and to participate in a machine learning colloquium called "The Importance of Being

Learnable." Some of the people who set up the foundations of learning theory (all of them CLRC Fellows) were among the invited speakers: Ray Solomonoff, one of the originators of algorithmic complexity, Chris Wallace, the creator of the Minimum Message Length approach to model selection, Jorma Rissanen, the creator of the Minimum Description Length principle, and Vladimir Vapnik and Alexey Chervonenkis.

Vladimir Vapnik gave a talk entitled "Learning Theory and Problems of Statistics," where he outlined methods for constructing a new type of universal learning machine described in his new book [5] on statistical learning published in the same year, 1998. Alexey's talk was devoted to the history of the Support Vector Machines. He reviewed developments over the last 30 years in the field of statistical learning theory and addressed the following topics:

- The "Generalized Portrait" as a minimax vector that characterizes a class in problems of pattern recognition.
- Converting the search for the "Generalized Portrait" to a convex programming problem.
- Support vectors and their properties implied by the Kuhn–Tucker theorem.
- Evaluation of generalization performance using the number of support vectors.

Alexey was appointed Professor of Computer Science at Royal Holloway in 2000 and then participated in many research projects at the CLRC. Among the most interesting ones were a combination of the Bayesian and maximum likelihood approaches to regularization, the development of a new method for recognition of promoter regions using string kernels, and exploring properties of systems with infinite VC dimension.

From 2010 Alexey was Emeritus Professor at Royal Holloway, University of London.

Yandex

In 2007 Alexey started lecturing at the School of Data Analysis founded by Yandex—the Russian equivalent of Google. The lectures were published as the book "Computer Data Analysis" [1]. Since 2011 he had also been working on a part-time basis at Yandex itself, developing machine learning algorithms related to the promotion of advertisements.

In 2012 Alexey published his new fundamental results about classes of events for which the conditions of uniform convergence are not satisfied. In 2013 and 2014 he published, with co-authors, research papers about his work at Yandex. In 2013 he presented two invited talks at big international conferences: the conference on data analysis organised by Yandex in Moscow and the conference "Artificial Intelligence Applications and Innovations." In 2013 and 2014 he was an honorary chair of the workshop "Conformal Prediction and Its Applications." He remained as active as ever.

Acknowledgment The editors are very grateful to the late Alexey Chervonenkis for his help and for sharing his recollections with them.

References

1. Chervonenkis, A.Y.: Компьютерный анализ данных (Computer Data Analysis, in Russian). Yandex, Moscow (2010)
2. Poznyak, A.S.: Воспоминания из другого тысячелетия (Recollections from another millennium, in Russian). In: Яков Залманович Цыпкин (1919–1997) (Yakov Zalmanovich Tsypkin (1919–1997)), pp. 130–156. Institute of Control Problems named after V.A. Trapeznikov, Moscow (2007)
3. Tamm, E.I.: Записки альпиниста (Alpinist's Notes, in Russian). ФИАН (Institute of Physics of the Russian Academy of Sciences), Moscow (2001)
4. Vapnik, V.N. (ed.): Алгоритмы и программы восстановления зависимостей (Algorithms and Programs for Dependency Reconstruction, in Russian). Nauka, Moscow (1984)
5. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
6. Vapnik, V.N., Chervonenkis, A.Y.: Об одном классе алгоритмов обучения распознаванию образов (On a class of algorithms for pattern recognition learning, in Russian, English summary). Автоматика и телемеханика (Automation and Remote Control) **25**(6), 937–945 (1964)
7. Vapnik, V.N., Chervonenkis, A.Y.: On the uniform convergence of the frequencies of occurrence of events to their probabilities. Doklady Akademii Nauk SSSR **181**, 781–783 (1968). Soviet Mathematics Doklady **9**, 915–918
8. Vapnik, V.N., Chervonenkis, A.Y.: On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and Its Applications **16**, 264–280 (1971). This volume, Chap. 3
9. Vapnik, V.N., Chervonenkis, A.Y.: Теория распознавания образов: Статистические проблемы обучения (Theory of Pattern Recognition: Statistical Problems of Learning: in Russian). Nauka, Moscow (1974). German translation: Theorie der Zeichenerkennung, transl. K.G. Stöckel and B. Schneider, ed. S. Unger and B. Fritzsch, Akademie Verlag, Berlin (1979)
10. Vapnik, V.N., Chervonenkis, A.Y.: Necessary and sufficient conditions for the uniform convergence of means to their expectations. Theory of Probability and Its Applications **26**(3), 532–553 (1982). Russian original: Теория вероятностей и ее применения **26**(3), 543–564 (1981)
11. Vapnik, V.N., Dronfort, L.M., Chervonenkis, A.Y.: Некоторые вопросы самоорганизации распозающих устройств (Some questions of the self-organization of recognizing systems, in Russian). In: Теория и применение автоматических систем (Theory and Application of Automatic Systems), pp. 172–177. Nauka, Moscow (1964)
12. Wiener, N.: Cybernetics: Or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, MA (1948). Russian translation: Винер, Н.: Кибернетика, или управление и связь в животном и машине. Советское радио, Moscow (1958)

Alexey's Colleagues

This section gives basic information about some of Alexey's colleagues mentioned in this book, including the members of the Aizerman–Braverman–Rozonoer group that was formed in 1960–1961 and carried out active research in pattern recognition.

The reader should keep in mind that there are two kinds of doctoral degrees in Russia: Ph.D. (кандидат наук) and the more advanced D.Sci. (доктор наук).

Mark Aronovich Aizerman (Марк Аронович Айзерман in Cyrillic)
Born in 1913 (Dvinsk, Russian Empire, nowadays Daugavpils, Latvia). A leading scientist in the area of control theory, one of the first cyberneticians in the USSR. Joined the Institute of Automation and Remote Control in 1939. Took part in WWII (1941–1945). Defended his D.Sci. thesis in 1946. Chair of the Department of Theoretical Mechanics in the Moscow Institute of Physics and Technology, where he worked part-time, between 1964 and 1978. Died in 1992.

Emmanuel Markovich Braverman (Эммануил Маркович Браверман in Cyrillic)
Born in 1931. Joined the Institute of Automation and Remote Control as Aizerman's Ph.D. student in 1960. Worked on the problem of pattern recognition. The author of the geometric approach to pattern recognition and the "compactness hypothesis" (briefly discussed in Chap. 5). Died in 1977.

Mikhail Moiseevich Bongard (Михаил Моисеевич Бонгард in Cyrillic)
Born in 1924 (Moscow). One of the pioneers of pattern recognition in the USSR. Since the early 1950s was a member of the laboratory of the biophysics of vision in the Institute of Biological Physics, which was transferred to the Institute of Information Transmission Problems and became the laboratory of information processing in sense organs in 1963; headed the laboratory in 1967–1971. Together with his colleagues developed the algorithm Kora (Russian abbreviation of "combinatorial recognition," комбинаторное распознавание), which was successfully applied to oil exploration. He was an outstanding mountaineer. Died in an accident in 1971 while descending a mountain; during the last break before his death told his team mates that because of the success of Kora in oil exploration his laboratory had been allowed to work in pure science and to choose the direction of its research ([3], p. 30).

Yakov Isaevich Khurgin (Яков Исаевич Хургин in Cyrillic)
Born in 1919 (Saratov, Russia). In 1960–1962 he was deputy head of the Research Council for Cybernetics of the USSR Academy of Sciences, and in 1962–1971 (the years when the VC dimension was born) he was Professor of the Department of Higher Mathematics of the Russian State University of Oil and Gas (modern name).

Aleksandr Yakovlevich Lerner (Александр Яковлевич Лернер in Cyrillic)
Born in 1913 (Vinnytsya, in the Pale of Jewish Settlement of Imperial Russia, now in Ukraine). A leading cyberneticist and since 1971 a prominent "refusenik" (a Soviet Jew wishing but not allowed to emigrate). In 1977, a letter in the Soviet newspaper "Izvestiya" called him "the leader of an espionage nest" (Wikipedia). There is evidence that Lerner's decision to emigrate led to a drastic change in Trapeznikov's attitude to anti-semitism in the early 1970s [2]. Emigrated to Israel in 1988. Died in 2004 (Rehovot, Israel).

Lev Il'ich Rozonoer (Лев Ильич Розоноер in Cyrillic)
Born in 1931. One of the founders of control theory. Joined the Institute of Automation and Remote Control in 1955. Part-time Professor in the Department of Theoretical Mechanics of the Moscow Institute of Physics and Technology from 1965. Moved to the USA in 1996.

Vadim Aleksandrovich Trapeznikov (Вадим Александрович Трапезников in Cyrillic)
Born in 1905 in Moscow. Studied at the same high school (E.A. Repman's gymnasium) as Andrei Kolmogorov. Graduated from the Bauman Moscow Higher Technical School (nowadays Bauman Moscow State Technical University) in 1928. Was awarded Ph.D. and D.Sci. degrees in the same year, 1938; the former was awarded using a special route, without submitting and defending a thesis. Joined the Institute of Automation and Remote Control in 1941, was its Director in 1951–1987. Died in 1994 in Moscow.

Vladimir Naumovich Vapnik (Владимир Наумович Вапник in Cyrillic)
Born in 1936 (Tashkent, Uzbekistan, USSR). Graduated from the Uzbek State University (now Samarkand State University) in 1958. Defended his Ph.D. thesis on the Generalized Portrait in 1964 at the Institute of Control Problems, and defended his D.Sci. thesis in 1984 (he defended the D.Sci. thesis with difficulty and only on his second attempt, mainly because of the anti-semitic attitudes at the ICS at the time; Chervonenkis was never awarded a D.Sci. degree and never submitted a D.Sci. thesis). From 1990 is based in the USA, his affiliations being AT&T Bell Labs (1990–2002) and NEC (2002–now); his other affiliations are Royal Holloway, University of London (1995–now) and Columbia University (2003–now).

Tragic Death of Alexey Chervonenkis

After the work on this book had been finished and the manuscript had been sent to the publishers, tragic news came from Russia: Alexey Chervonenkis went for a walk in a large national park on the outskirts of Moscow, lost his way at night, and died of hypothermia. This happened on 22 September 2014, just a few weeks after Alexey's 76th birthday. Alexey's tragic death was reported in leading Russian and British newspapers, and obituaries by his colleagues at the University of London and Yandex appeared on the Web straight away [1–4]. They said, in particular, that Alexey's contribution to the theory of machine learning was enormous and that he was an exceptional teacher and a great friend.

Here is the brief story. On Sunday, 21 September 2014, at 14:00 (Moscow time) Alexey left his home for a long walk in Elk Island (Лосиный Остров), a huge (116 km^2) and beautiful national park divided between Moscow and the Moscow region. Alexey loved long walks and hikes, and in his younger years often walked for 40–50 miles. In recent years distances went down to 15–20 miles, but walking remained a regular activity, at least 2–3 times a week. He recorded all his walks, and in 2014 he made 158 in Moscow (plus many more in England during his visit in August 2014). It was a very warm day, with a temperature reaching 20 °C, and he dressed lightly: a suit and light jumper, casual low shoes, no hat. The map (Fig. 1) shows his intended route, but nobody knows what route he actually took (Fig. 2).

At 20:20 he called his wife Vika to say that he had got lost in the forest, was tired, and wet after falling into a swamp, but would try to get out by himself. Several more calls followed, but Alexey insisted that he would be able to manage the situation.

At about 23:00 when Vika called him again, he repeated that he was very tired and his clothes were wet, his cigarette lighter was wet and did not work—and all he wanted was to have some rest and then to resume his way out of the forest.

The last call came around 0:30; Alexey said that he wanted to take a nap to get some strength and asked Vika not to disturb him for a few hours. After that the line went dead. The temperature dropped sharply, to −2 °C towards the morning.

Fig. 1 The intended route of Alexey's last walk in Elk Island (moving broadly North, which is upwards on the map). The Moscow Automobile Ring Road is marked by its Russian abbreviation МКАД. The large blue comma marks the place where he died

Fig. 2 Elk Island as painted by Alexey Savrasov in 1869 (*left*) and a modern photo (*right*)

In the morning of 22 September Vika went to Moscow and alerted their three sons. They immediately called the police, the State Rescue Service, and other organisations in Moscow and in the area of the park. After long, painful, and ultimately useless negotiations with the state bureaucracy, real help came from a voluntary organisation called Liza Alert. They gathered about 20 people and organized the search. By the evening of the 22nd, when the news of Alexey's disappearance spread, many colleagues and students joined the search—overall, there were over 180 people, who divided the park into sectors and spent all night searching. It was a difficult search, at night in treacherous swamps with many small ponds, and finding a person lying in high grass was nearly impossible. The people were doing their best, but all they found were eyeglasses that belonged to Alexey.

On the next day, 23 September, Alexey's eldest son Michael hired a helicopter and flew to the volunteers' headquarters over the park. He realized that finding a person from high altitude was very difficult and he could easily miss Alexey. Luckily, there was a professional among the volunteers, a marksman, who flew to

the place where the glasses had been discovered and almost immediately spotted a body lying in a clearing in the forest. However, it was impossible to land the helicopter there. This happened around 12:00. Soon a second, bigger, helicopter came and managed to land in a nearby clearing. The marksman discovered that it was a dead man, took a picture of the body, and Michael identified his father.

As a forensic examination showed later, Alexey died from hypothermia. The drop in temperature from 22 °C in daytime to −2 °C at night appeared to be crucial. In difficult topographical conditions, exacerbated by poor equipment and bureaucratic delays, it took rescuers, police, and medical services around ten hours to retrieve Alexey's body from the swamp. He was buried on 25 September 2014 at the Khovansky cemetery in Moscow.

Michael mentioned that when he eventually managed to get to the place, he saw that Alexey was lying on his back with a happy expression on his face. Apparently, people who die from hypothermia go to sleep and feel a relief as though they reach their destination.

Rest in peace, dear colleague and friend.

Acknowledgments Many thanks to Michael Chervonenkis, Alexey's son, for providing much of the information about the events surrounding Alexey's death. Our other sources of information include Liza Alert's website, Russian and British newspapers, and Russian Wikipedia.

References

1. Johnston, I.: Renowned mathematician dies after getting lost during walk in Moscow woods. The Independent (26 September 2014)
2. Luhn, A.: University of London maths professor found dead in Moscow park. Guardian (25 September 2014)
3. Tribute to Alexey Chervonenkis. Royal Holloway, University of London (24 September 2014). URL https://www.royalholloway.ac.uk/
4. Volozh, A.: In memory of Alexey Yakovlevich Chervonenkis (Памяти Алексея Яковлевича Червоненкиса). Yandex blog (29 September 2014). URL http://blog.yandex.ru/post/88529

Contents

Contributors

Luca Baldassarre LIONS, EPFL, Lausanne, Switzerland

Peter L. Bartlett Departments of Electrical Engineering and Computer Science and Statistics, UC Berkeley, Berkeley, CA, USA; School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

Léon Bottou Microsoft Research, New York, NY, USA

Olivier Catoni CNRS – UMR 8553, Département de Mathématiques et Applications, École Normale Supérieure, Paris, France; INRIA Paris-Rocquencourt – CLASSIC Team, Le Chesnay Cedex, France

Vladimir Cherkassky Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA

Alexey Chervonenkis Institute of Control Sciences, Moscow, Russia; Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, UK; Yandex, Moscow, Russia

Koby Crammer Technion – Israel Institute of Technology, Haifa, Israel

Sauptik Dhar Research and Technology Center, Robert Bosch LLC, Palo Alto, CA, USA

R.M. Dudley Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Ran El-Yaniv Technion – Israel Institute of Technology, Haifa, Israel

László Györfi Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest, Hungary

Dominik Janzing Max Planck Institute for Intelligent Systems, Tübingen, Germany

Yuri Kalnishkan Department of Computer Science and Computer Learning Research Centre, Royal Holloway, University of London, Egham, Surrey, UK

Andreas Maurer Munich, Germany

Vasily N. Novoseltsev Institute of Control Sciences, Moscow, Russia

Asaf Noy Technion – Israel Institute of Technology, Haifa, Israel

Massimiliano Pontil University College London, London, UK

Alexander Rakhlin University of Pennsylvania, Philadelphia, PA, USA

Benjamin I.P. Rubinstein Department of Computing and Information Systems, The University of Melbourne, Melbourne, Australia

J. Hyam Rubinstein Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia

Bernhard Schölkopf Max Planck Institute for Intelligent Systems, Tübingen, Germany

Naji Shajarisales Max Planck Institute for Intelligent Systems, Tübingen, Germany

Alexander Shen LIRMM UM2 Montpellier, Montpellier, France; Institute for Information Transmission Problems, Moscow, Russia

Karthik Sridharan University of Pennsylvania, Philadelphia, PA, USA

Ingo Steinwart Institute for Stochastics and Applications, University of Stuttgart, Stuttgart, Germany

Bastian Steudel Max Planck Institute for Intelligent Systems, Tübingen, Germany

V.N. Vapnik NEC Laboratories America, Princeton, NJ, USA

Nikolay Vereshchagin Moscow State University and Yandex, Moscow, Russia

Vladimir Vovk Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, UK

Vladimir V. V'yugin Institute for Information Transmission Problems, Moscow, Russia

Harro Walk Department of Mathematics, University of Stuttgart, Stuttgart, Germany

Yair Wiener Technion – Israel Institute of Technology, Haifa, Israel

Acronyms

ALBA	Active-Learning-Based Algorithm
CAL	Cohn–Atlas–Ladner (in "CAL algorithm")
CDF	Cumulative distribution function
CLRC	Computer Learning Research Centre
CPSU	Communist Party of the Soviet Union
CSS	Consistent selective strategy
ERM	Empirical Risk Minimization
GP	Generalized Portrait
IAT	Institute of Automation and Remote Control
ICS	Institute of Control Sciences
IGCI	Information-Geometric Causal Inference
i.i.d.	Independent and identically distributed
KL	Kullback–Leibler (in "KL divergence")
LDA	Linear discriminant analysis
MAD	Median absolute deviation
MIPT	Moscow Institute of Physics and Technology
MLP	Multilayer perceptron
PAC	Probably approximately correct
PCA	Principal component analysis
PDF	Probability density function
QDA	Quadratic discriminant analysis
RBF	(as in RBF kernel) Radial basis function
RKHS	Reproducing kernel Hilbert space
ROC	Receiver operating characteristic
RV	Random variable
SLLN	Strong law of large numbers
SOM	Self-organizing Map
SSL	Semisupervised learning

SVM	Support vector machine
T-SVM	Transductive SVM
TL	Transductive learning
TM	Turing machine
VC	Vapnik–Chervonenkis (as in "VC bounds," "VC classes," "VC dimension," "VC theory," etc.)

Credits

- The photograph in the frontmatter is used with permission from Alexey Chervonenkis.
- The photographs in the introductions to Parts II and IV are used with permission from Valentina Fedorova.
- The following English translation by B. Seckler of the 1971 paper by Vladimir Vapnik and Alexey Chervonenkis is included with permission of the copyright holder, SIAM.
  V.N. Vapnik and A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities (translated by B. Seckler). Theory of Probability and its Applications, Vol. 16, Number 2, pp. 264–280, 1971 © 1971 by SIAM.

Part I
History of VC Theory

Part I of the book is devoted to the history of VC theory. It opens (Chap. 1) with an English translation of a unique historical document written by Alexey Chervonenkis in Russian in 2004 and entitled "Воспоминания Червоненкиса" (Chervonenkis's recollections). This 20-page handwritten manuscript was not intended for publication and was circulated privately among a limited number of Alexey's colleagues. It is a very open account of the early years of the collaboration between Vladimir Vapnik and Alexey Chervonenkis up to the publication in 1974 of their fundamental monograph [5]. Another set of Alexey Chervonenkis's reminiscences was published in 2013 in the Vapnik Festschrift ([2], Chap. 3).

Chapter 3 is a reprint of the famous 1971 paper by Vapnik and Chervonenkis introducing fundamental notions and proving fundamental results of VC theory. The short 1968 paper announcing those results was reprinted in the Vapnik Festschrift ([2], Chap. 2). The significance of the 1971 paper is explained in Richard Dudley's introduction, Chap. 2.

The following chapter (Chap. 4) by Richard Dudley is based on his historical talk at the "Measures of Complexity" symposium in Cyprus. It contains an engaging review of the "ancient" history of VC theory, whose elements can be traced back to Steiner (1826) and Schläfli (1901). The crucial result, however, is Vapnik and Chervonenkis's dichotomy: either the growth function is exponential, $n \mapsto 2^n$ (the trivial case), or bounded by a polynomial. The history of this result and the closely related notion of VC dimension are also briefly covered by Dudley's review.

Léon Bottou's contributions to this volume (Chaps. 9 and 12) are about sharpening VC bounds in statistical learning theory, but his talk (Fig. III.1 in the introduction to Part III) at the "Measures of Complexity" symposium was in fact devoted to a different topic, namely the discovery of various versions of Vapnik and Chervonenkis's dichotomy. Its title was "About the origins of the Vapnik–Chervonenkis Lemma," and it gave further details of the history of this discovery. It appears that Vapnik and Chervonenkis were the first to come up with the dichotomy, and their papers and their reviews by Dudley (Sect. 4.8) intrigued famous mathematicians such as Paul Erdös and via them diffused into the work of other

mathematicians, including Sauer [3] and Shelah [4]. Nowadays, the name "Sauer's lemma" is often used to refer to the optimal version of the dichotomy first proved by Sauer [3]; see Sect. 4.5 for details. In their 1974 monograph [5] Vapnik and Chervonenkis also proved the optimal version. For further information on the history of Vapnik and Chervonenkis's dichotomy, see Bottou's paper [1] in the Vapnik Festschrift and Chap. 5 in this volume.

The last chapter in this part, Chap. 5, is a fascinating account of Vladimir Vapnik and Alexey Chervonenkis's work at the Institute of Control Sciences in the 1960s and 1970s. It is written by Vasily Novoseltsev, Vladimir and Alexey's close colleague, and is full of funny, interesting, and instructive details (sometimes these qualities are even combined). Some of the highlights are the Soviet system for approving foreign trips for scientists, Alexey's role in proving Vapnik and Chervonenkis's dichotomy, and Alexey's correspondence with Michel Talagrand.

Added in proofs: Sadly, Vasily Novoseltsev died on 24 March 2015, before the publication of this book. He was born in 1935 in Arkhangelsk, Russia, studied at the MIPT in 1953–1959, and for the rest of his life worked at the Institute of Control Sciences.

References

1. Bottou, L.: In hindsight: Doklady Akademii Nauk SSSR, 18(4), 1968. In: Schölkopf B., Luo Z., Vovk V. (eds.) Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, pp. 3–5. Springer, Berlin (2013)
2. Chervonenkis, A.: Early history of support vector machines. In: Schölkopf B., Luo Z., Vovk V. (eds.) Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, pp. 13–20. Springer, Berlin (2013)
3. Sauer, N.: On the density of families of sets. J. Comb. Theor. Series A **13**, 145–147 (1972)
4. Shelah, S.: A combinatorial problem: stability and order for models and theories in infinitary languages. Pac. J. Math. **41**, 247–261 (1972)
5. Vapnik, V.N., Chervonenkis, A.Y.: Теория распознавания образов: Статистические проблемы обучения (Theory of Pattern Recognition: Statistical Problems of Learning: in Russian). Nauka, Moscow (1974). German translation: Theorie der Zeichenerkennung, transl. K.G. Stöckel and B. Schneider, ed. S. Unger and B. Fritzsch, Akademie Verlag, Berlin (1979)

Chapter 1
Chervonenkis's Recollections

Alexey Chervonenkis

Abstract These recollections about the origins of VC theory were written by Alexey Chervonenkis in 2004 for several colleagues and not intended for publication. They are now published for the first time. (Eds.) Translated by Vladimir Vovk.

The original setting of the problem (Autumn 1962) of pattern recognition learning was as follows. There are N decision rules (ways of dividing objects into classes). The teacher is using one of them. A training sequence $x_1, \ldots, x_l$ is given, and the teacher classifies it naming for each point its class $\omega_1, \ldots, \omega_l$ using one of the N known rules. The learning machine excludes from the list those rules that make an error, i.e., work *differently* from the teacher. There remain N_1 rules, and they are bound to contain the true one. (We called such algorithms *algorithms with complete memory*, as opposed to recurrent ones.)

The idea was to prove that there exists a training sequence such that N_1 becomes equal to 1, i.e., *only the true decision rule remains*, and, moreover, $l \sim \log N$. This scheme is reminiscent of searching for a counterfeit coin using a series of weighings (in which case $l \sim \log N$ is indeed sufficient).

An almost inverse statement is easy to show for recognition of binary vectors. If we want each of the N decision rules to be chosen given some training sequence, the number N must be at most the number of all variants of the training sequence of length l. For binary vectors of dimension n the number of such variants is equal to $2^{(n+1)l}$. From this we get

A. Chervonenkis
Institute of Control Sciences, Laboratory 38, Profsoyuznaya Ulitsa 65,
Moscow 117997, Russia

A. Chervonenkis
Department of Computer Science, Royal Holloway, University of London,
Egham, Surrey, UK

A. Chervonenkis
Yandex, Moscow, Russia

V. Vovk et al. (eds.), *Measures of Complexity*,
DOI 10.1007/978-3-319-21852-6_1

$$N \leq 2^{(n+1)l},$$
$$l \geq \frac{\log_2 N}{n+1}.$$

Therefore the length of the sample must be at least[1] $\frac{\log_2 N}{n+1}$.

It turned out, however, that the opposite inequality

$$l \lesssim \ln N$$

in this setting is, in general, not correct.

For example, let there be $N-1$ objects, $N-1$ decision rules each of which assigns one of the objects to class I and the rest to class II, and the Nth decision rule assigning all objects to class II. If the teacher is using the last rule, all the given objects will be assigned to class II and only at most l decision rules will be discarded, and to discard all of them (except for the right one) everything has to be shown. That is, $l = N - 1$ rather than $\log N$.

Up to this point the problem did not involve *probability*. The indicated failure, and also other considerations, forced us to change the setting in March 1963.

The training sequence is not *given* but *generated* by some source Γ independently with a constant, but unknown, distribution $P(x)$ (the i.i.d. hypothesis). On the other hand, we do not require that only one decision rule remains in reserve, but allow arbitrarily many provided they are *close* to the true one, i.e., make an error with probability $< \varkappa$ (under the same distribution as for training). Then it is easy to get a logarithmic estimate.

The probability that a rule that is different from the true one by more than $\varkappa$ will not be eliminated on a sample of length l is less than

$$p = (1 - \varkappa)^l.$$

The probability that at least one such rule will not be eliminated is less than

$$N(1 - \varkappa)^l.$$

[1] In fact, in the setting of the problem as described here it is also true that

$$N \leq 2^l$$
$$l \geq \log_2 N$$

($\log_2$ standing for base 2 logarithm). Alexey's weaker (but sufficient for his purpose) bound $(\log_2 N)/(n+1)$ also holds in a situation that is easier for the learner: he knows the true decision rule, and his goal is to choose a training sequence $x_1, \ldots, x_l$ proving that the known decision rule is indeed the true one (in the sense that the observed $\omega_1, \ldots, \omega_l$ is compatible with only one rule). (Eds.)