

Stroje a množiny Zdzisława Pawlaka

Jiří Novotný

Ústav matematiky a deskriptivní geometrie

Stavební fakulta VUT v Brně

Žižkova 17, 662 37 Brno

e-mail: novotny.j@fce.vutbr.cz

Abstrakt

Příspěvek je věnován vynikajícímu představiteli matematických základů informatiky Zdzisławu Pawlakovi, který letos v dubnu zemřel v nedožitých osmdesáti letech. Hlavní pozornost je upřena na tzv. Pawlakovy stroje – model programovatelného počítače a na teorii hrubých množin – nástroj pro analýzu vágních a nepřesných dat.

Profesor Zdzisław Pawlak se narodil 10. listopadu 1926 v Lodži, umřel 7. dubna 2006 ve Varšavě.



Z. Pawlak získal světovou proslulost, když v roce 1982 zformuloval novou teorii, představující doplnění teorie množin, kterou nazval teorií hrubých množin (anglicky rough set theory). Tématům spojeným s touto teorií se věnovalo mnoho mezinárodních konferencí, bylo publikováno několik tisíc vědeckých článků a vydáno několik desítek knih. Z. Pawlak je též znám ve spojitosti s konstrukcí prvních polských elektronických počítačů GAM-1 v padesátých letech minulého století. Navrhl novou metodu generování náhodných čísel, kterou opublikoval v roce 1953 v prestižním americkém vědeckém časopise. Byla to pravděpodobně první vědecká práce z informatiky polského autora uveřejněná v zahraničí. Dále se Z. Pawlak věnoval minusdvojkové aritmetice. Pod jeho vedením byl sestaven počítačový stroj UMC 1 pracující podle tohoto teoretického modelu.

Posléze se profesor Pawlak zabýval teoretickou informatikou. Speciálně navrhl novou třídu bezkontextových jazyků zobecňující známou polskou bezkontextovou notaci, představil nový formální model počítače nazývaný potom jeho jménem – Pawlakův stroj (anglicky Pawlak machine). Pracoval také na formálním matematickém modelu genetických kódů DNA a přinesl nový matematický přístup v teorii konfliktů.

Profesní a kariérní vývoj Z. Pawlaka začal v Lodži, kde v roce 1939 ukončil základní školu. V době okupace během druhé světové války pracoval ve firmě Siemens. Po válce v roce 1946 udělal externě maturitu a v roce 1947 nastoupil na fakultu elektrotechniky Politechniky v Lodži, odkud přešel na fakultu spojovací techniky Politechniky ve Varšavě. Diplom magistra inženýra spojů získal v roce 1951. Po skončení studia pracoval do roku 1957 v Institutu matematiky Polské akademie věd jako asistent. V letech 1957–1959 pracoval na Varšavské Politechnice. Přitom na Institutu základních problémů techniky Polské akademie věd získal v roce 1958 doktorát (kandidát technických věd) za práci Použití teorie grafů v syntéze překladačů. Docentem (habilitovaným doktorem) v oboru matematika se stal v roce 1963 po obhajobě práce Struktura bezadresových strojů. Titul mimořádného profesora získal v Institutu matematiky Polské akademie věd v roce 1971. Řádným profesorem matematiky byl jmenován v Institutu základů

informatiky Polské akademie věd v roce 1978, kde byl v letech 1971–1979 zástupcem ředitele pro vědeckou činnost. V roce 1983 se stal členem korespondentem Polské akademie věd, řádným členem v roce 1991. Od roku 1985 až do roku 2006 pracoval v Institutu teoretické i aplikované informatiky Polské akademie věd v Glivicích. V letech 1989–1996 byl ředitelem Institutu informatiky na fakultě elektroniky Politechniky ve Varšavě. A konečně v letech 1998–2006 působil i na Vyšší škole aplikované informatiky a zařízení.

Profesor Z. Pawlak vydal několik knih a více než 200 vědeckých článků v renomovaných nakladatelstvích s mezinárodním dosahem. Byl školitelem asi 30 aspirantů (doktorandů) z oblasti matematiky a informatiky. Byl více než stokrát pozván na mezinárodní konference, semináře a další akce, aby proslovil referáty na témata těchto vědeckých setkání. Několikrát působil jako hostující profesor na univerzitách ve Spojených státech, Kanadě a Evropě, kde přednášel matematické základy informatiky, matematickou logiku, strukturu číselných počítačů, matematickou lingvistiku a jiné.

Z. Pawlak získal mnoho cen a vyznamenání, státních i vědeckých, byl členem asi 20 vědeckých rad, mnoha redakcí národních i zahraničních časopisů, mj. byl zástupcem hlavního redaktora Bulletinu Polské akademie věd. Z jeho iniciativy byl založen známý časopis *Fundamenta Informaticae*.

Měl jsem tu čest profesora Pawlaka osobně poznat, hovořit s ním na seminářích na brněnské univerzitě a spolupracovat s ním ve dvou oblastech jeho širokého vědeckého záběru a to v problematice Pawlakových strojů a hrubých množin. Proto se v další části své práce budu tímto oblastem podrobněji věnovat.

V roce 1969 Z. Pawlak prezentoval v práci *Maszyny programowane* (Programovatelné stroje) zjednodušený teoretický model imitující aktivity programovatelného počítače. Je to vlastně množina stavů s parciální unární operací. Jestliže začneme v nějakém stavu, program buď směřuje výpočet do dalšího stavu, nebo výpočet ukončí. Tedy z algebraického hlediska *Pawlakův stroj* je uspořádaná trojice (A, f, r) , kde A je množina, f parciální unární operace na A a r unární relace na A taková, že $r = A - \text{dom } f$. Je-li dán stav $x \in A$, pak stroj přechází do stavu $f(x)$, pak do stavu $f^2(x)$, atd. Buď tato posloupnost nazývaná výpočtem je nekonečná, nebo v případě, že $f^n(x) \notin \text{dom } f$ pro nějaké $n \geq 0$, je konečná. Homomorfismy a simulace těchto strojů byly definovány a studovány Wiktorem Bartolem v roce 1974 v práci *Programy dynamiczne obliczeń* (Dynamické programy výpočtů). Protože teorie Pawlakových strojů je blízká teorii monounárních algeber, někteří brněnští matematici se snažili přenést své metody a výsledky z úplných monounárních algeber na parciální monounární algebry a Pawlakovy stroje. Jedná se např. o práci Miroslava Novotného *On some problems concerning Pawlak's machines* z roku 1975 a *On mappings of machines* z roku 1976, Oldřicha Kopečka *Construction of all machine homomorphisms* z roku 1976 a *Homomorphisms of machines* z roku 1978 a Jiřího Novotného *The category of Pawlak machines* z roku 1982 a Další vlastnosti kategorie Pawlakových strojů z roku 1990. Tyto odkazy demonstrují, že práce *Maszyny programowane* profesora Z. Pawlaka byla velmi inspirující a přispěla mj. k prokázání užitečnosti studia monounárních algeber.

Nejvýznamnější oblastí, ve které Z. Pawlak působil, jak už jsem uvedl, je problematika hrubých množin a jejich aplikací. Pojem hrubé množiny Pawlak představil v práci *Rough Sets* v roce 1982 jako matematický nástroj pro analýzu vágních a nepřesných dat. Tento přístup je originální a liší se od podobné teorie fuzzy množin Lofti A. Zadeha z roku 1965. Ukázalo se, že Pawlakem navržený přístup je důležitý v mnoha disciplínách imitace intelektu (anglicky *artificial intelligence*), jako např. ve strojovém učení, rozpoznávání obrazců, analýze dat, v rozhodovacích a expertních systémech apod.

Výchozím bodem Pawlakových úvah je zjištění, že v mnoha aplikacích je dána množina objektů (stavů, procesů, pozorování atd.), ale my je nejsme schopni rozlišit pomocí dostupných měření, pozorování nebo popisů. Jinými slovy, náš nedostatek znalostí může být vážnou překážkou při rozhodování o objektech, fenoménech, procesech atd. Tedy základní poznatek je, že

vágnost nebo nepřesnost vede k nerozlišitelnosti, která formálně reprezentuje náš nedostatek znalostí a je základní myšlenkou podtrhující filozofii hrubých množin. Relace nerozlišitelnosti se používá k definování základních operací na množinách – dolní a horní aproximaci množiny, které se využívají místo přesných pojmů. Obvykle ztotožňujeme pojmy s podmnožinami nějakého univerza. S každou množinou dat pak nakládáme jako se speciálním druhem rozhodovací tabulky a dolní a horní aproximace se používají k analýze vlastností rozhodovacích tabulek. Např. rozhodovací tabulka může obsahovat data o pacientech trpících nějakou určitou chorobou. Symptomy pacientů jsou pak považovány za podmínkové atributy a např. zdravotní stav pacienta se dá považovat za rozhodovací atribut. Pak hlavní otázkou je, zda zdravotní stav pacienta se dá definovat pomocí jeho symptomů. Jinými slovy, zajímáme se o závislosti mezi atributy v rozhodovací tabulce. Přístup teorie hrubých množin se ukázal pro analýzu otázek tohoto typu jako velmi vhodný. V současnosti byla implementována a úspěšně aplikována celá řada systémů analýzy počítačových dat, a to v průmyslu, lékařské sociologii, psychologii, pedagogice a dalších oborech.

Matematicky vycházíme z pojmu *aproximační prostor* jako dvojice $A = (U, R)$, kde U je množina nazývaná univerzum a R je binární relace na U nazývaná *relace nerozlišitelnosti*. Tato relace musí být samozřejmě reflexivní a symetrická, tj. $R(x, x)$ a $R(x, y)$ implikuje $R(y, x)$ pro všechna $x, y \in U$. Tedy v obecném případě jde o tzv. relaci tolerance. Běžně se však předpokládá, že relace nerozlišitelnosti je také tranzitivní, tj. $R(x, y)$ a $R(y, z)$ implikuje $R(x, z)$ pro všechna $x, y, z \in U$. Tedy se předpokládá, že R je relace ekvivalence. Tento předpoklad je motivován mnoha praktickými aplikacemi uvedeného pojmu. Jestliže R je relace nerozlišitelnosti, pak R^* označuje soubor všech tříd ekvivalence relace R (rozkladů U podle R .) Tyto třídy ekvivalence se nazývají *R-elementární množiny* aproximačního prostoru A . R-elementární množinu obsahující prvek $x \in U$ označíme $[x]_R$. Každé sjednocení R-elementárních množin se nazývá *R-rozlišitelná množina*. Jestliže množina není R-rozlišitelná, budeme ji nazývat *hrubá množina* v aproximačním prostoru A .

Relace nerozlišitelnosti určitým způsobem reprezentuje náš nedostatek znalostí o prvcích univerza. Čím máme více znalostí, tím jemnější je odpovídající aproximační prostor a v důsledku toho jsme schopni přesněji rozlišovat prvky univerza.

Základní roli hraje pojem aproximace množiny jinou množinou. Předpokládejme, že je dán aproximační prostor $A = (U, R)$ a podmnožina $X \subseteq U$. Definujeme dvě množiny

$$\underline{R}X = \{x \in U; [x]_R \subseteq X\} \quad \text{a} \quad \overline{R}X = \{x \in U; [x]_R \cap X \neq \emptyset\}$$

nazývané *R-dolní* a *R-horní aproximace množiny X*.

Množina $BN_R(X) = \overline{R}X - \underline{R}X$ se nazývá *R-hraniční obor množiny X*. Také se používá označení $POS_R(X) = \underline{R}X$, tzv. *R-pozitivní obor množiny X*, $NEG_R(X) = U - \overline{R}X$, tzv. *R-negativní obor množiny X*.

Pozitivní a negativní obor představují množiny všech prvků univerza U , které se dají správně klasifikovat do množin X a $-X$ v aproximačním prostoru A .

Číslo

$$\alpha_R(X) = \frac{\text{card } \underline{R}X}{\text{card } \overline{R}X}$$

se nazývá *R-přesnost množiny X*.

Číslo $\rho_R(X) = 1 - \alpha_R(X)$ se označuje za *R-hrubost množiny X*.

Zřejmě platí $\alpha_R(X) \geq 0$, $\rho_R(X) \leq 1$. Snadno se ukáže platnost následujících vlastností.

a) X je R-rozlišitelná množina, právě když $\underline{R}X = \overline{R}X$.

b) X je hrubá vzhledem k R , právě když $\underline{R}X \neq \overline{R}X$.

Dále si povšimněme, že $\underline{R}X$ je maximální R -rozlišitelná množina obsažená v množině X a $\overline{R}X$ je minimální R -rozlišitelná množina obsahující množinu X . Poznamenejme, že pojem aproximace množiny vede k následujícím definicím dvou relací příslušnosti definovaných vztahy

$$x \in_R X, \text{ právě když } x \in \underline{R}X \quad \text{a} \quad x \in \overline{R}X, \text{ právě když } x \in \overline{R}X,$$

kde symbol $\in_R$ čteme „ x určitě patří do X vzhledem k R “ a $\in \overline{R}$ „ x možná patří do X vzhledem k R “.

Z topologického hlediska se hrubé množiny dají klasifikovat následujícím způsobem.

- a) Jestliže $\underline{R}X \neq \emptyset$ a $\overline{R}X \neq U$, pak množina X je *hrubě R -rozlišitelná*.
- b) Jestliže $\underline{R}X = \emptyset$ a $\overline{R}X \neq U$, pak množina X je *interně R -nerozlišitelná*.
- c) Jestliže $\underline{R}X \neq \emptyset$ a $\overline{R}X = U$, pak množina X je *externě R -nerozlišitelná*.
- d) Jestliže $\underline{R}X = \emptyset$ a $\overline{R}X = U$, pak množina X je *totálně R -nerozlišitelná*.

Do dalších podrobností teorie hrubých množin se v této práci nebudeme pouštět. Na závěr přikládám seznam některých anglicky psaných prací profesora Z. Pawlaka věnovaných této problematice.

2005

- [57] Zdzisław Pawlak: Rough Sets and Flow Graphs. RSFDGrC (1) 2005: 1-11
- [56] Zdzisław Pawlak: Flow Graphs and Data Mining. T. Rough Sets 2005: 1-36
- [55] Zdzisław Pawlak, Lech Polkowski, Andrzej Skowron: Rough Sets. Encyclopedia of Database Technologies and Applications 2005: 575-580
- [54] Zdzisław Pawlak: A Treatise on Rough Sets. T. Rough Sets: 1-17 (2005)

2004

- [53] Zdzisław Pawlak: Inference Rules and Decision Rules. ICAISC 2004: 102-108
- [52] Zdzisław Pawlak: Decision Networks. Rough Sets and Current Trends in Computing 2004: 1-7
- [51] Salvatore Greco, Zdzisław Pawlak, Roman Slowinski: Bayesian Confirmation Measures within Rough Set Approach. Rough Sets and Current Trends in Computing 2004: 264-273
- [50] Zdzisław Pawlak: Some Issues on Rough Sets. T. Rough Sets 2004: 1-58
- [49] Zdzisław Pawlak: Decision Rules and Dependencies. Fundam. Inform. 60(1-4): 33-39 (2004)

2003

- [48] Zdzisław Pawlak: Flow Graphs and Decision Algorithms. RSFDGrC 2003: 1-10
- [47] Zdzisław Pawlak: Probability, Truth and Flow Graph. Electr. Notes Theor. Comput. Sci. 82(4): (2003)
- [46] Zdzisław Pawlak: A rough set view on Bayes' theorem. Int. J. Intell. Syst. 18(5): 487-498 (2003)

2002

- [45] Zdzislaw Pawlak: The Rough Set View on Bayes' Theorem. AFSS 2002: 106-120
- [44] Zdzislaw Pawlak: Theorize with Data Using Rough Sets. COMPSAC 2002: 1125-1128
- [43] James F. Peters, Zdzislaw Pawlak, Andrzej Skowron: A Rough Set Approach to Measuring Information Granules. COMPSAC 2002: 1135-1139
- [42] Zdzislaw Pawlak: In Pursuit of Patterns in Data Reasoning from Data The Rough Set Way. Rough Sets and Current Trends in Computing 2002: 1-9
- [41] Salvatore Greco, Zdzislaw Pawlak, Roman Slowinski: Generalized Decision Algorithms, Rough Inference Rules, and Flow Graphs. Rough Sets and Current Trends in Computing 2002: 93-104
- [40] Zdzislaw Pawlak: Rough sets and intelligent data analysis. Inf. Sci. 147(1-4): 1-12 (2002)

2001

- [39] Zdzislaw Pawlak: Bayes' Theorem Revised - The Rough Set View. JSAI Workshops 2001: 240-250
- [38] Zdzislaw Pawlak, James F. Peters, Andrzej Skowron, Zbigniew Suraj, Sheela Ramanna, Maciej Borkowski: Rough Measures and Integrals: A Brief Introduction. JSAI Workshops 2001: 375-379
- [37] Zdzislaw Pawlak: Combining Rough Sets and Bayes' Rules. Computational Intelligence 17(3): 401-408 (2001)
- [36] Zdzislaw Pawlak: Drawing conclusions from data - The rough set way. Int. J. Intell. Syst. 16(1): 3-11 (2001)
- [35] Zdzislaw Pawlak: Rough Set Theory. KI 15(3): 38-39 (2001)

2000

- [34] Zdzislaw Pawlak: Rough Sets and Decision Algorithms. Rough Sets and Current Trends in Computing 2000: 30-45

1999

- [33] Zdzislaw Pawlak: Logic, Probability, and Rough Sets. Jewels are Forever 1999: 364-373
- [32] Zdzislaw Pawlak: Data Mining - a Rough Set Perspective. PAKDD 1999: 3-12
- [31] Zdzislaw Pawlak: Decision Rules, Bayes' Rule and Ruogh Sets. RSFDGrC 1999: 1-9
- [30] Zdzislaw Pawlak: Rough classification. Int. J. Hum.-Comput. Stud. 51(2): 369-383 (1999)

1998

- [29] Zdzislaw Pawlak: Reasoning about Data - A Rough Set Perspective. Rough Sets and Current Trends in Computing 1998: 25-34
- [28] Zdzislaw Pawlak: An Inquiry into Anatomy of Conflicts. Inf. Sci. 109(1-4): 65-78 (1998)

1997

- [27] Zdzislaw Pawlak: Vagueness - a Rough Set View. Structures in Logic and Computer Science 1997: 106-117

1996

- [26] Zdzislaw Pawlak: Rough Sets, Rough Relations and Rough Functions. Fundam. Inform. 27(2/3): 103-108 (1996)

1995

- [25] Zdzislaw Pawlak: Rough Sets. ACM Conference on Computer Science 1995: 262-264
- [24] Zdzislaw Pawlak, Jerzy W. Grzymala-Busse, Roman Slowinski, Wojciech Ziarko: Rough Sets. Commun. ACM 38(11): 88-95 (1995)
- [23] Zdzislaw Pawlak: Vagueness and Uncertainty: A Rough Set Perspective. Computational Intelligence 11: 277-232 (1995)

1993

- [22] Zdzislaw Pawlak: Hard and Soft Sets. RSKD 1993: 130-135
- [21] Zdzislaw Pawlak: Knowledge and Uncertainty: A Rough Set Approach. SOFTEKS Workshop on Incompleteness and Uncertainty in Information Systems 1993: 34-42
- [20] Zdzislaw Pawlak: Anatomy of confics. Bulletin of the EATCS 50: 234-246 (1993)

1992

- [19] Zdzislaw Pawlak: Concurrent versus sequential - the rough sets perspective. Bulletin of the EATCS 48: 178-190 (1992)
- [18] Miroslav Novotný, Zdzislaw Pawlak: On a problem concerning dependence spaces. Fundam. Inform. 16(3-4): 275-287 (1992)

1991

- [17] Zdzislaw Pawlak: Decision Logik. Bulletin of the EATCS 44: 201-225 (1991)
- [16] Miroslav Novotný, Zdzislaw Pawlak: Algebraic theory of independence in information systems. Fundam. Inform. 14(4): 454-476 (1991)

1990

- [15] Zdzislaw Pawlak: Theory of Rough Sets: A New Methodology for Knowledge Discovery (Abstract). ICCI 1990: 11

1989

- [14] Zdzislaw Pawlak: Knowledge, reasoning and classification: A rough set perspective. Bulletin of the EATCS 38: 199-210 (1989)

1988

- [13] Zdzislaw Pawlak, S. K. Michael Wong, Wojciech Ziarko: Rough Sets: Probabilistic versus Deterministic Approach. International Journal of Man-Machine Studies 29(1): 81-95 (1988)

1987

- [12] Zdzislaw Pawlak: Decision tables - a rough set approach. Bulletin of the EATCS 33: 85-95 (1987)

1986

- [11] Zdzislaw Pawlak, Krzysztof Slowinski, Roman Slowinski: Rough Classification of Patients After Highly Selective Vagotomy for Duodenal Ulcer. International Journal of Man-Machine Studies 24(5): 413-433 (1986)

1984

- [10] Zdzislaw Pawlak: On rough sets. Bulletin of the EATCS 24: 94-108 (1984)
- [9] Zdzislaw Pawlak: Rough Classification. International Journal of Man-Machine Studies 20(5): 469-483 (1984)
- [8] Ewa Orlowska, Zdzislaw Pawlak: Expressive Power of Knowledge Representation Systems. International Journal of Man-Machine Studies 20(5): 485-500 (1984)
- [7] Zdzislaw Pawlak: On Conflicts. International Journal of Man-Machine Studies 21(2): 127-134 (1984)
- [6] Ewa Orlowska, Zdzislaw Pawlak: Representation of Nondeterministic Information. Theor. Comput. Sci. 29: 27-39 (1984)

1981

- [5] Zdzislaw Pawlak: Information systems theoretical foundations. Inf. Syst. 6(3): 205-218 (1981)

1979

- [4] Zdzislaw Pawlak, Grzegorz Rozenberg, Walter J. Savitch: Programs for Instruction Machines. Information and Control 41(1): 9-28 (1979)

1976

- [3] V. Wiktor Marek, Zdzislaw Pawlak: Information Storage and Retrieval Systems: Mathematical Foundations. Theor. Comput. Sci. 1(4): 331-354 (1976)

1973

- [2] Zdzislaw Pawlak: A mathematical model of digital computers. Automatentheorie und Formale Sprachen 1973: 16-22
- [1] Zdzislaw Pawlak: Mathematical Foundation of Information Retrieval. MFCS 1973: 135-136