

Již tradičně byly příspěvky rozděleny do několika sekcí – původní sekce „TeX a priatelia“ byla rozšířena na „TeX, R a ich priatelia“, v níž dominoval neúnavný přispěvatel a organizátor Pavel Striž. Významnou se stala sekce věnující se IoT a otevřenému hardwaru a také tradičně jsme měli možnost naslouchat zajímavým příspěvkům sekce zabývající se Open GIS a Open Data. Doprovodnou akcí byl workshop *Missing Maps mapathon Žilina #14*.

Prínosným byl také paralelní workshop *Žnázornenie a interpretácia dát v marimo notebooku* s charismatickým Michalem Kaučičem (předchozí hlavní organizátor).

Příspěvky nezařazené do těchto témat pak tvořily vhodnou výplň, v níž se začíná profilovat všudypřítomná umělá inteligence. V diskusi jsme se pokoušeli alespoň částečně v tomto objemném tématu vymezit, jak AI ovlivňuje výuku, softwarová řešení a další aspekty pedagogické a vědecké práce.

Účastníci měli možnost se potkat na neformálním posezení u výběrného guláše a pěnívého moku, ale nejen tam si mohli vyměnit spoustu zajímavých informací, zkušeností a námětů.

Velký dík patří organizátorům celé akce – Aleši Kozubíkovi, Pavlu Strižovi, Rudolfu Blaškovi, Miloslavu Ofukanému i všem dalším, kteří přispěli ke zdárnému průběhu celé akce. Věřme, že příští OSSConf se podaří zorganizovat podobně, přestože v realizačním týmu dojde ke změně – hlavní organizátor Aleš Kozubík symbolicky předal žezlo Miroslavu Kvašayovi.

Jiří Rybicka

LINEÁRNÍ REGRESNÍ MODEL FUZZY ČASOVÉ ŘADY LINEAR REGRESSION MODEL OF FUZZY TIME SERIES

Zdeněk Karpíšek¹, Veronika Lacinová², Jakub Šácha³,
Tomaš Pospíšil⁴

Adresa: ^{1,4}Odbor statistiky a optimalizace, Ústav matematiky, Fakulta strojního inženýrství, Vysoké učení technické v Brně, Technická 2, 616 69 Brno,
²Katedra kvantitativních metod, Fakulta vojenského leadershipu, Univerzita obrany v Brně, Kounicova 156/65, 662 10 Brno,
³Ústav statistiky a operační analýzy, Provozně ekonomická fakulta, Mendelova univerzita v Brně, Zemědělská 1, 613 00 Brno.

E-mail: ¹karpisek@fme.vutbr.cz, ²veronika.lacinova@unob.cz,

³jakub.sacha@mendelu.cz, ⁴tomin.pospiseznam.cz

Abstrakt: V příspěvku je pomocí fuzzy aritmetiky popsán vliv nepřesností na stanovení lineární aproximace trendu a periodicity vývoje ekonomického ukazatele vyjádřeného časovou řadou. Je prezentováno spojení základního aparátu lineární regresní analýzy, fuzzy aritmetiky a jeho užití pro výpočty odhadů fuzzy hodnot regresních koeficientů a fuzzy hodnot pozorované časové řady.

Klíčová slova: časová řada, lineární regresní funkce, fuzzy časová řada, fuzzy lineární regresní funkce, odhady.

Abstract: This paper describes using fuzzy arithmetic an influence of inaccuracies to determine a linear approximation of trend and periodicity of an economic indicator expressed by a time series. The combination of the basic apparatus of linear regression analysis and fuzzy arithmetic is presented and its use for the calculation of estimates of fuzzy values of regression coefficients and fuzzy values of observed time series.

Keywords: time series, linear regression function, fuzzy time series, fuzzy linear regression function, estimates.

1. Úvod

Motivací k tomuto příspěvku bylo získat odhady trendové a periodické složky časové řady z expertních nebo statistických intervalových odhadů hodnot pozorované ekonomické veličiny (znaku, ukazatele), jejíž dynamiku vyvoje tato

časová řada popisuje, a to pomocí regresní analýzy a fuzzy aritmetiky. Důvodem k fuzzy pojetí je skutečnost, že se v praxi obvykle hodnoty časových ekonomických indikátorů berou jako zcela přesné a nerespektuje se fakt, že jsou nepřesné z důvodu působení např. inflace na cenu měnové jednotky, nepřesnosti měření apod. Závěry o aproximaci a predikci dané časové řady z těchto „nefuzzy“ hodnot proto nemusi odpovídat skutečnosti. Tento příspěvek je rozšířením výsledků popsaných v [2] a [3], kde jde o vyjádření neurčitosti pozorovaných hodnot pomocí intervalové analýzy. Nové získané výsledky pomocí fuzzy modelování jsou popsány v Oddílu 3 a 4 tohoto příspěvku.

2. Lineární regresní model časové řady

Znamé pojmy a jejich vlastnosti uvedené v tomto oddílu je možno najít např. v [1], [6], [7].

Časová řada je posloupnost reálných čísel $\mathbf{x} = (x_1, \dots, x_n)$, kde $x_i = x(t_i)$ je pozorovaná hodnota náhodné veličiny X_i v čase t_i , $t_i < t_{i+1}$, $i = 1, \dots, n - 1$ a $n > 2$.

Lineární regresní funkce má tvar $x(t) = \sum_{j=1}^m \beta_j f_j(t)$, kde reálné funkce $f_j(t)$ reálné proměnné t neobsahují neznámé parametry a β_j , $j = 1, \dots, m$, jsou **regresní koeficienty**.

Lineární regresní model časové řady $\mathbf{x} = (x_1, \dots, x_n)$ je založen na předpokladu, že hodnoty časové řady jsou hodnoty náhodných veličin $X_i = \sum_{j=1}^m \beta_j f_j(t_i) + E_i$, $m < n$. Pro jednoduchost předpokládáme, že všechny náhodné veličiny E_i mají nulovou střední hodnotu, stejný rozptyl a jsou nekorelované, $i = 1, \dots, n$.

Lineární regresní funkce $x(t) = \sum_{j=1}^m \beta_j f_j(t)$ vyjadřuje obvykle trendovou složku dané časové řady $\mathbf{x} = (x_1, \dots, x_n)$, ale může také současně zahrnovat její periodickou složku.

Odhad hodnoty časové řady v čase t pomocí regresní analýzy je funkce $\hat{x}(t) = \sum_{j=1}^m b_j f_j(t)$, kde b_j , $j = 1, \dots, m$, jsou **odhady regresních koeficientů** β_j .

Pro odhady použijeme tyto matice:

$$\mathbf{F} = \begin{pmatrix} f_1(t_1) & f_1(t_2) & \cdots & f_1(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(t_1) & f_m(t_2) & \cdots & f_m(t_n) \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix},$$



Miloslaw Ojlikaný uvádí sekci Open GIS/Open Data



Aleš Kozubík odměňuje symbolickým dárkem Michala Kouckého



Zvaná přednáška Ondřeje Vencálka



Zvaná přednáška Slavka Fedorika

$$\mathbf{G} = \mathbf{F}\mathbf{F}^T = \begin{pmatrix} \sum_{i=1}^n f_1^2(t_1) & \sum_{i=1}^n f_1(t_i)f_2(t_i) & \cdots & \sum_{i=1}^n f_1(t_i)f_m(t_i) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n f_m(t_i)f_1(t_i) & \sum_{i=1}^n f_m(t_i)f_2(t_i) & \cdots & \sum_{i=1}^n f_m^2(t_i) \end{pmatrix}$$

$$= \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix},$$

$$\mathbf{g} = \mathbf{F}\mathbf{x}^T = \begin{pmatrix} \sum_{i=1}^n f_1(t_i)x(t_i) \\ \vdots \\ \sum_{i=1}^n f_m(t_i)x(t_i) \end{pmatrix} = \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix},$$

$$\mathbf{G}^{-1} = \begin{pmatrix} g^{11} & g^{12} & \cdots & g^{1m} \\ \vdots & \vdots & \ddots & \vdots \\ g^{m1} & g^{m2} & \cdots & g^{mm} \end{pmatrix}, \quad \mathbf{f}(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_m(t) \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

kde horní index T značí transponovanou matici.

Dále předpokládáme, že matice $\mathbf{F}$ má hodnot m , takže matice $\mathbf{G}$ je regulární a inverzní matice $\mathbf{G}^{-1}$ existuje. Následující odhady jsou získány tzv. **metodou nejmenších čtverců**.

Odhady regresních koeficientů b_j , $j = 1, \dots, m$, jsou řešením soustavy normálních rovnic $\mathbf{G}\mathbf{b} = \mathbf{g}$, takže

$$\mathbf{b} = \mathbf{G}^{-1}\mathbf{g} = \mathbf{G}^{-1}\mathbf{F}\mathbf{x}^T$$

a vidíme, že odhady b_j , $j = 1, \dots, m$, jsou lineárními funkcemi hodnot $x_i = x(t_i)$ pozorované časové řady. Použitím uvedených matic dostaneme po úpravách $b_j = \sum_{i=1}^n c_{ji}x_i$, kde pro $j = 1, \dots, m$ je $c_{ji} = \sum_{k=1}^m g^{jk}f_{ki}$. Odhad hodnoty časové řady v čase t je

$$\hat{x}(t) = \mathbf{f}^T(t)\mathbf{b} = \mathbf{f}^T(t)\mathbf{G}^{-1}\mathbf{g} = \mathbf{f}^T(t)\mathbf{G}^{-1}\mathbf{F}\mathbf{x}^T.$$

Tento odhad je zřejmě lineární funkcí hodnot pozorované časové řady $\mathbf{x} = (x_1, \dots, x_n)$. Po úpravách dostaneme

$$\hat{x}(t) = \sum_{i=1}^n a_i(t)x_i,$$

kde pro $i = 1, \dots, n$ je

$$a_i(t) = \mathbf{f}^T(t) \mathbf{G}^{-1} \mathbf{F} = \sum_{j=1}^m f_j(t) \sum_{k=1}^m g^{jk} f_{ki} = \sum_{j=1}^m c_{ji} f_j(t).$$

3. Fuzzy čísla – základní pojmy a vlastnosti

Modelování nepřesných hodnot reálných čísel pomocí fuzzy množin [4] a [5], tj. jejich *fuzifikace*, vychází z následujících pojmů.

Fuzzy množina $\tilde{x} = (\mu_{\tilde{x}}; \mathbb{R})$ se nazývá (*reálné*) *fuzzy číslo*, jestliže tato fuzzy množina je konvexní a normální, a funkce příslušnosti $\mu_{\tilde{x}} : \mathbb{R} \rightarrow [0; 1]$ je po částech spojitá. Jestliže existuje právě jedno číslo $x \in \mathbb{R}$ takové, že $\mu_{\tilde{x}}(x) = 1$, pak x je *hlavní hodnota* fuzzy čísla $\tilde{x}$. Je-li funkce příslušnosti $\mu_{\tilde{x}}$ spojitá, je $\tilde{x}$ *spojité fuzzy číslo*. Množinu všech fuzzy čísel na $\mathbb{R}$ značíme $\mathfrak{F}$. Klademe $x = \{x\}$ pro $\forall x \in \mathbb{R}$, takže každé reálné číslo x je také fuzzy číslo. Podobně jako reálné fuzzy číslo definujeme *celé fuzzy číslo* apod.

Unární a binární operace s reálnými čísly fuzifikujeme, tj. zavádíme *rozšířené unární* a *binární operace* s fuzzy čísly pomocí tzv. *Zadehova principu rozšíření*.

Jestliže $\varphi(x)$ je unární operace na $\mathbb{R}$, pak *rozšířená unární operace* $\varphi(\tilde{x})$ na $\mathfrak{F}$ je fuzzy množina $\tilde{y}$ s funkcí příslušnosti

$$\mu_{\tilde{y}}(y) = \sup_{\varphi(x)=y} \mu_{\tilde{x}}(x).$$

Speciálně pro bijekci $\varphi(x)$ je $\mu_{\tilde{y}}(y) = \mu_{\tilde{x}}(\varphi^{-1}(y))$.

Jestliže je $x \star y$ binární operace na $\mathbb{R}$, pak *rozšířená binární operace* $\tilde{x} \star \tilde{y}$ na $\mathfrak{F}$ je fuzzy množina $\tilde{z}$ s funkcí příslušnosti

$$\mu_{\tilde{z}}(z) = \sup_{\substack{x, y \\ x \star y = z}} \left\{ \mu_{\tilde{x}}(x), \mu_{\tilde{y}}(y) \right\}.$$

Jestliže $\star$ je binární operace na $\mathbb{R}$, pak v $\mathfrak{F}$ platí:

- (a) $\star$ komutativní $\Rightarrow \star$ komutativní,
- (b) $\star$ asociativní $\Rightarrow \star$ asociativní.

Rozšířené aritmetické operace $\oplus, \otimes, \ominus, \odot$ s fuzzy čísly mají řadu analogických vlastností jako aritmetické operace $+, \cdot, -, /$ včetně jejich priorit. Pro

ZPRÁVA O KONFERENCI OSSCONF 2024

REPORT ON THE OSSCONF 2024 CONFERENCE

Jiří Rybíčka

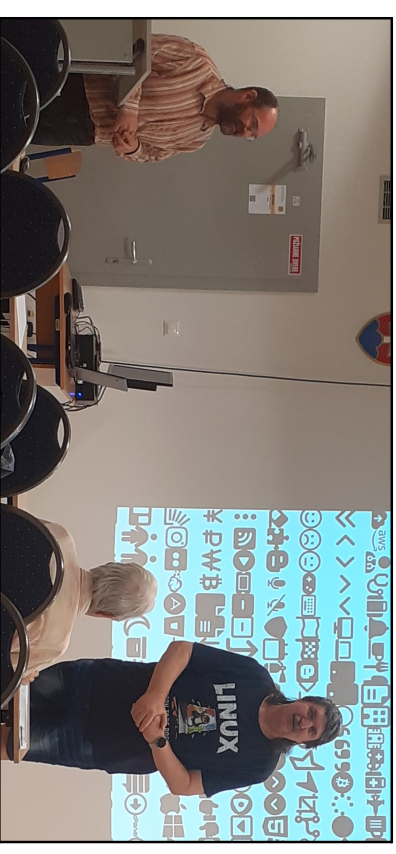
E-mail: rybi.cka@mendelu.cz

OSSConf 2024

S radostí a určitým očekáváním jsem přijel do Žiliny, abych se účastnil 12. ročníku konference *Otvorený softvér vo výuke, výskumu a IT riešeniach*. Tentokrát jsme se scházeli po pětileté přestávce způsobené různými okolnostmi, o to raději jsem se potkal se „starými známými“ i novými tvářemi.

Podobně jako v minulých letech se organizátoři snažili, aby konference byla přístupná všem – bez vložného, s minimálními náklady na sborník a ostatní výdaje. O to srdečnější a „domáctější“ je atmosféra konference.

Zvané přednášky byly tři a všechny zaujaly přítomné posluchače. První na řadě byla *Sudoku s překryvy* (Pavel Stríž), druhá pak *Vizuálizace dat o úmrtnosti v Anglii pomocí softwaru R a interpretace srovnání měř úmrtnosti u očkovaných a neočkovovaných proti onemocnění COVID-19* (Ondřej Vencálek) a v posledním dnu pak třetí – *Ansbile: Automatizácia správų systémů a softwaru* (Slavko Fedorik).



Aleš Kozubík uvádí zvanou přednášku Pavla Stríže

Kapitola 8: More Advanced Topics. Kapitola zmiňuje těžké úlohy dynamického programování, průtoky sítí, ale také úlohy kategorie NP-úplné. Poslední podkapitola se věnuje rozložení těžkých úloh na menší, z pohledu algoritmu na samostatně řešitelné.

Kapitola 9: Rare Topics. Vůči dřívějším vydáním této knihy zbrusu nová kapitola. Zmiňuje atypické a speciální úlohy. Došlo i na Billa Gatese, neb jeho jediný odborný článek spadá k úloze Pancake Sorting, viz str. 551–553. Zaujala mě kapitola 9.30, tedy konstrukční úlohy: vytvoření magického čtverce pro velké n , podobně jako úloha rozmístění N královen na šachovnici. Za zvláštní pozornost stojí úlohy interaktivní s elektronickým systémem (kapitola 9.31), setkal jsem se s tím u úloh na bioinformatických soutěžích.

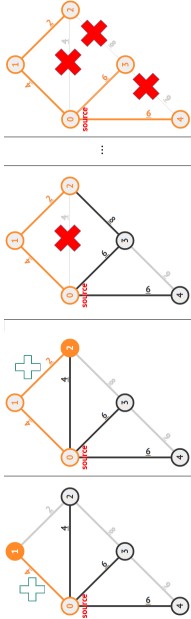
Za velmi důležitou kapitolu považuji na str. 590 tu poslední (9.34 Chapter Notes), tedy co autoři chtěli v knize mít, ale nestihli to. To lze považovat za inspiraci k dalšímu studiu nad rámcem této knihy.

Ze zajímavých knižních projektů, vhodné jako studijní pokračování pro adepty soutěžního programování, bych rád zmínil:

- Antti Laaksonen: *Guide to Competitive Programming*, Springer, 2017. Web: <https://cses.fi/problemset/>.
- Jörg Arndt: *Matters Computational. Ideas, Algorithms, Source Code* (alias fxtbook), 2010. Web: <https://www.jjj.de/fxt>.
- Joseph O'Rourke: *Computational Geometry in C*, Cambridge University Press, 2nd edition, 1998. Web: <https://www.science.smith.edu/~jorourke>.

Každá z knih obsahuje seznam zdrojů, rejstřík i medailónky tvůrců algoritmů. Je vysázena v $\text{\LaTeX}$. Kniha se nedá číst jako román, ale neměla by chybět v (e-)knižovničce programátora jako užitečná referenční příručka.

Nasínice určité potěši zmínění Vojtěcha Jarníka (byť tedy „ě“ je špatně vysázené jako „ê“) v první knize na str. 217 a 222, a na Michala Foriška, viz zdroje [16] a [17] na str. 274 první knihy a [15] na str. 591 druhé knihy. Václav Chvátal nesměl chybět, viz druhá kniha, str. 282, Art Gallery Problem. Kdysi jsem mu psal email kolem jeho příručky k $\text{\LaTeX}$ u, krásná vzpomínka. Na počest Vojtěcha Jarníka zmiňuji ukázkou animace ze str. 218 první knihy.



realizaci rozšířených operací $\oplus, \otimes, \ominus, \oslash$ pomocí α -řezů spojitých fuzzy čísel lze použít tzv. *intervalovou aritmetiku* [3]. Při fuzzifikaci „klasických“ číselných matematických modelů vytvářením adekvátních fuzzy modelů se používají vybrané třídy expertně stanovených fuzzy čísel. Dostatečně flexibilním typem fuzzy čísel s ohraničeným nosičem a jednoprvkovým jádrem, tj. hlavní hodnotou, jsou pro naše účely spojitá fuzzy čísla definovaná následujícím způsobem.

Nechť $\Lambda : [0; 1] \rightarrow [0; 1]$ je spojitá rostoucí funkce taková, že $\Lambda(0) = 0$ a $\Lambda(1) = 1$. **Λ -fuzzy číslem** $\underline{x} = (x_l, x_m, x_r)_\Lambda$, kde $x_l, x_m, x_r \in \mathbb{R}$ a $x_l < x_m < x_r$, rozumíme fuzzy číslo s funkcí příslušnosti

$$\mu_{\underline{x}}(x) = \begin{cases} \Lambda\left(\frac{x - x_l}{x_m - x_l}\right), & x \in [x_l, x_m], \\ \Lambda\left(\frac{x_r - x}{x_r - x_m}\right), & x \in [x_m, x_r], \\ 0, & \text{jinde.} \end{cases}$$

Funkce Λ se nazývá **generátor** fuzzy čísla $\underline{x} = (x_l, x_m, x_r)_\Lambda$ a interval (x_l, x_r) je **nosič** tohoto fuzzy čísla. Platí:

1. Jestliže $\underline{x} = (x_l, x_m, x_r)_\Lambda$ a $c \in \mathbb{R}$, pak **rozšířený násobek**

$$c\underline{x} = \begin{cases} (cx_l, cx_m, cx_r)_\Lambda, & \text{když } c > 0, \\ (cx_r, cx_m, cx_l)_\Lambda, & \text{když } c < 0, \\ 0, & \text{když } c = 0. \end{cases}$$

2. Jestliže $\underline{x} = (x_l, x_m, x_r)_\Lambda$ a $\underline{y} = (y_l, y_m, y_r)_\Lambda$, pak **rozšířený součet**

$$\underline{x} \oplus \underline{y} = (x_l + y_l, x_m + y_m, x_r + y_r)_\Lambda.$$

3. Jestliže $\underline{x}_i = (x_l, x_m, x_r)_\Lambda$ a $a_i \in \mathbb{R}$, $i = 1, \dots, n$, pak **rozšířená lineární kombinace**

$$\begin{aligned} \sum_{i=1}^n a_i \tilde{x}_i &= \left(\sum_{\substack{i=1 \\ a_i > 0}}^n a_i x_{il} + \sum_{\substack{i=1 \\ a_i < 0}}^n a_i x_{ir}, \sum_{i=1}^n a_i x_{im}, \sum_{\substack{i=1 \\ a_i > 0}}^n a_i x_{ir} + \sum_{\substack{i=1 \\ a_i < 0}}^n a_i x_{il} \right)_{\Lambda} \\ &= \left(\sum_{i=1}^n a_i \left(\frac{1 + \operatorname{sgn}(a_i)}{2} x_{il} + \frac{1 - \operatorname{sgn}(a_i)}{2} x_{ir} \right), \sum_{i=1}^n a_i x_{im}, \right. \\ &\quad \left. \sum_{i=1}^n a_i \left(\frac{1 + \operatorname{sgn}(a_i)}{2} x_{ir} + \frac{1 - \operatorname{sgn}(a_i)}{2} x_{il} \right) \right)_{\Lambda} \\ &= \left(\sum_{i=1}^n a_i \frac{x_{il} + x_{ir}}{2} - \sum_{i=1}^n |a_i| \frac{x_{ir} - x_{il}}{2}, \sum_{i=1}^n a_i x_{im}, \right. \\ &\quad \left. \sum_{i=1}^n a_i \frac{x_{il} + x_{ir}}{2} + \sum_{i=1}^n |a_i| \frac{x_{ir} - x_{il}}{2} \right)_{\Lambda}. \end{aligned}$$

4. Fuzzy lineární regresní model fuzzy časové řady

Jestliže místo náhodných hodnot $x_i = x(t_i)$ časové řady $\mathbf{x} = (x_1, \dots, x_n)$ uvažujeme fuzzy hodnoty $\tilde{x}_i$, tj. fuzzy čísla, jejichž nosiče obsahují x_i , $i = 1, \dots, n$, dostaneme **fuzzy časovou řadu** $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$. Původní časovou řadu pozorovaných hodnot x_i při praktických aplikacích fuzzyfikujeme tak, že tyto hodnoty nahradíme fuzzy čísla $\tilde{x}_i$ a to obvykle expertně.

Fuzzy lineární regresní funkce má tvar

$$\tilde{x}(t) = \sum_{j=1}^m \tilde{\beta}_j f_j(t),$$

kde funkce $f_j(t)$ neobsahují neznámé parametry a $\tilde{\beta}_j$, $j = 1, \dots, m$, jsou **fuzzy regresní koeficienty**.

Fuzzy lineární regresní model fuzzy časové řady $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ je založený na předpokladu, že hodnoty fuzzy časové řady jsou fuzzy náhodné veličiny stejně jako v „nefuzzy“ případě.

Spojením výsledků z Oddílů 2 a 3 získáme snadno následující **fuzzy odhady**.

Odhad fuzzy regresního koeficientu $\tilde{\beta}_j$, $j = 1, \dots, m$, je fuzzy číslo

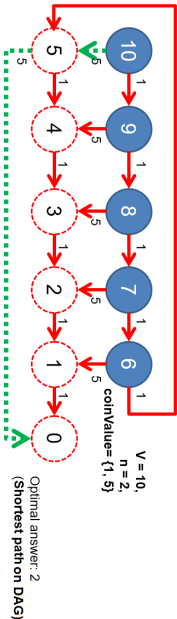
$$\tilde{b}_j = \sum_{i=1}^n c_{ji} \tilde{x}_i,$$

Kapitola 1: Introduction. Obsahuje typy a základní praktiky soutěžního programování, které se liší od běžného programování, např. načtení všech STL knihoven C++ místo vybraných (`bits/stdc++.h`), užívání zkratk a konstant a též `using namespace std`;

Kapitola 2: Data Structures and Libraries. Kapitola představuje lineární a nelineární datové struktury. Velká část kapitol se věnuje binárním rozhodovacím stromům – Self-balancing Binary Search Tree (AVL, RB). U C++ se zde poprvé setkáváme s načtením struktur mimo základní STL (`bits/extc++.h`). Co mě trochu zaskočilo, že podrobně komentují vlastní datové struktury v tak brzké fázi knihy.

Kapitola 3: Problem Solving Paradigms. Tato kapitola uvádí základní strategie řešení úloh: Complete Search, Divide and Conquer, Greedy Approach a Dynamic Programming.

Kapitola 4: Graph. Kapitola je úvod do teorie grafů, byt' tedy rozsáhlá. Zaujalo mě užít upravené verze Dijkstrova algoritmu a převedení úlohy mincí (Coin-Change) na hledání nejkratších cest v orientovaném acyklickém grafu (zkr. DAG), viz str. 253 a 254. Zde je grafická ukázka.



V poznámce pod čarou je odkaz na článek jednoho z autorů, kde pracoval s velkými grafy se 411 milióny vrcholů a 31 miliardami hran.

Kapitola 5: Mathematics. Tato kapitola se hlouběji zabývá úlohami matematiky (teorie čísel, práce s maticemi), kombinatoriky a pravděpodobnosti. Jsou zde zmíněné odkazy na <https://projecteuler.net> a <https://brilliant.org>, které obsahují nespočet úloh se zaměřením na matematiku. Speciálně na první projekt je stále častěji odkazováno v souvislosti se zajímavými úlohami.

Kapitola 6: String Processing. Práce s textovými řetězci je náplní této kapitoly. Na str. 364 zmiňuje algoritmy, které v knize nemají, ale považují je za důležité.

Kapitola 7: (Computational) Geometry. Kapitola se zaměřuje na geometrii 0D až 3D, s tím, že pro 3D je úloha nejméně, jak na soutěžích nebývají typické. Autoři zmiňují, že adaptovali algoritmy ze serveru <https://shygy.py.com/tools/>.

RECENZE KNIHY: COMPETITIVE PROGRAMMING 4 BOOK REVIEW: COMPETITIVE PROGRAMMING 4

Pavel Stríž

E-mail: pavel@striz.cz

Steven Halim, Felix Halim, Suhendry Effendy: Competitive Programming 4, Book 1 (xxx + 1-299 pp., tedy 329 stran PDF), Book 2 (viii + 273-616 pp., tedy 352 stran PDF).

Motto knihy, repeat for a lifetime:
Study; Practice; Rehearse; Dress Rehearse; Perform.

Knih k soutěžnímu programování najdeme řadu, všechny hezké (souhrn viz str. 50 první knihy), zde zmíněná kniha patří k nejnovějším. Nepoužívá pseudo-algoritmy, až na výjimky, ale ukazuje přímá řešení. Obsahuje komentované příklady, neřešené příklady s řešením na konci kapitol a také těžší příklady bez řešení, uváděné s hvězdičkou.

Cílem knihy je připravit na dva typy soutěží, IOI (The International Olympiad in Informatics, <https://ioinformatics.org>), to je spíš náplní první knihy a ICPC (The International Collegiate Programming Contest, <https://icpc.global>), to je rozšiřující náplní druhé knihy. První soutěž je do 18 let, pro jednotlivce, druhá soutěž je pro týmy vysokoškoláků. Co mě zaujalo, že na IOI opouští Pascal, ale i C. To mi u C přijde divné.

Kniha se odkazuje na zadání na serverech UVa (nově Online Judge; <https://uhunt.onlinejudge.org/>) a Kattis (<https://open.kattis.com>). Příklady obsahují *Entry level*, tedy doporučený první příklad kategorie; *Must try* *, tedy takové, které by si čtenář měl zkusit a pak ostatní doporučené příklady, řekněme bonusové. Autoři dohromady vyřešili 3458 úloh z těchto serverů, a typy, jak je vyřešit, zveřejnili na <https://cpbook.net/methodstosolve>.

Řešené příklady, častokrát pro 4 programovací jazyky, lze nalézt na GitHubu jednoho z autorů (<https://github.com/stevenhalim/cpbook-code>), jazyky jsou C++, Python, Java a OCaml. Co mě tolik nepřekvapilo je, že na soutěžích ubývá příkladů s velkými čísly.

Podpůrné stránky knih jsou <https://cpbook.net> a na vizualizaci algoritmů připravili server <https://visualgo.net>.

Velmi stručné představím jednotlivé kapitoly.

kde $c_{ji} = \sum_{k=1}^m g^{jk} f_{kl}$.

Odhad fuzzy hodnoty fuzzy časové řady v čase t je fuzzy číslo

$$\hat{x}(t) = \sum_{j=1}^m b_j f_j(t) = \sum_{i=1}^n a_i(t) \underline{x}_i,$$

kde pro $i = 1, \dots, n$ je $a_i(t) = \sum_{j=1}^m c_{ji} f_j(t)$.

Poznamenejme, že uvedené fuzzy odhady jsou sice „bodové“, ale mají charakter intervalových odhadů podobně jako při aplikaci lineárních stochastických regresních modelů.

Jestliže všechny nosiče fuzzy čísel $\underline{x}_i$ jsou ohraničené intervaly, pak je také nosič odhadu fuzzy hodnoty $\hat{x}(t)$ ohraničený interval a jeho délka je tzv. **šířka** $\delta(t)$ odhadu fuzzy hodnoty fuzzy časové řady v čase t . Pro fuzzy hodnoty $\underline{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$ pozorované časové řady je

$$\delta(t) = \sum_{i=1}^n a_i(t) \operatorname{sgn}(a_i(t)) (x_{ir} - x_{il}).$$

Fuzzifikaci pozorované časové řady $\mathbf{x} = (x_1, \dots, x_n)$ provádíme ve dvou krocích:

1. Zvolíme expertně tvar funkce příslušnosti Λ –fuzzy čísel $\underline{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$ pro fuzzifikaci hodnot x_i časové řady výběrem vhodného generátoru Λ pro $i = 1, \dots, n$.
2. Určíme hlavní hodnoty Λ –fuzzy čísel tak, že položíme $x_{im} = x_i$ a expertně stanovíme krajní body nosičů x_{il}, x_{ir} Λ –fuzzy čísel $\underline{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$ pro $i = 1, \dots, n$.

5. Příklad užití fuzzy lineárního regresního modelu

Pro jednoduchost ilustrujeme předcházející výsledky na „školním“ příkladu neperiodické časové řady s lineárním trendem. V tomto případě má regresní funkce tvar

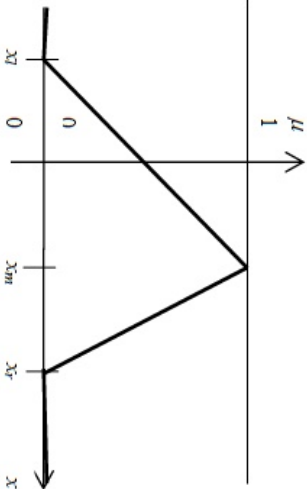
$$x(t) = \beta_1 + \beta_2 t,$$

takže $m = 2$, $f_1(t) = 1$ a $f_2(t) = t$. Pro fuzzifikaci časové řady používáme bez újmny na obecnosti nejčastěji aplikovaná **trojúhelníková fuzzy čísla**

$\tilde{x} = (x_l, x_m, x_r)_\Lambda$, jejichž generátor je $\Lambda(x) = x$, takže mají funkci příslušnosti

$$\mu_{\tilde{x}}(x) = \begin{cases} \frac{x - x_l}{x_m - x_l}, & x \in [x_l, x_m], \\ \frac{x_r - x}{x_r - x_m}, & x \in [x_m, x_r], \\ 0, & \text{jinde.} \end{cases}$$

Graf funkce příslušnosti trojúhelníkového fuzzy čísla je na Obrázku 1.



Obrázek 1: Graf funkce příslušnosti trojúhelníkového fuzzy čísla.

Zadání původní časové řady $\mathbf{x} = (x_1, \dots, x_n)$ a fuzzy časové řady $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ po expertní fuzifikaci nahrazením hodnot $x_i = x(t_i)$ fuzzy čísla $\tilde{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$, kde klademe $x_{im} = x_i$, je pro $n = 4$ v Tabulce 1. Jednoduchým výpočtem pomocí vzorců z Oddílů 2 a 4 dostaneme odhady fuzzy regresních koeficientů $\tilde{\beta}_1, \tilde{\beta}_2$:

$$\begin{aligned} \tilde{b}_1 &= (b_{1l}, b_{1m}, b_{1r})_\Lambda = (0; 1,35; 2,125)_\Lambda, \\ \tilde{b}_2 &= (b_{2l}, b_{2m}, b_{2r})_\Lambda = (-0,125; 0,24; 0,725)_\Lambda. \end{aligned}$$

Odhady hodnot fuzzy časové řady $\tilde{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$ a odhady hodnot $\hat{x}(t)$ ve zvolených časech t jsou v Tabulce 2. Graf

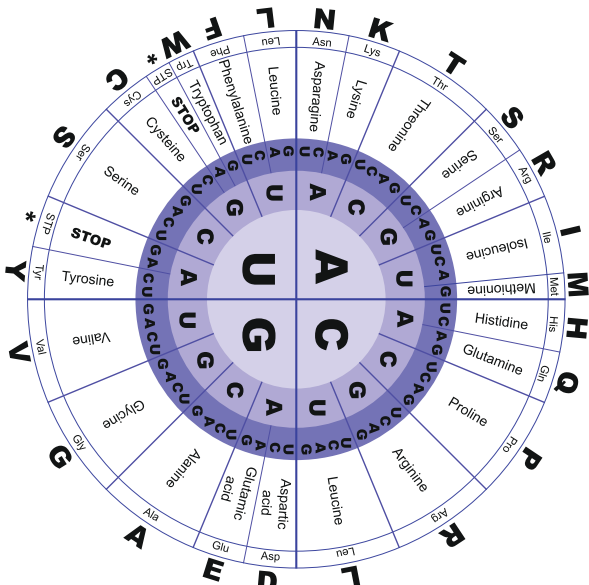
- nosičů (svíslé tečkované čáry) fuzzy časové řady $\tilde{x}_i = (x_{il}, x_{im}, x_{ir})_\Lambda$,
- hranic nosičů a hlavních hodnot (spojité čáry) odhadů hodnot fuzzy časové řady,

nejlepší soutěžní programátor), aby si užil hezkou knihu a „trochu“ toho programování.

Knihla obsahuje hezké ilustrace, rozšiřující seznam zdrojů a našeince potěší, že mezi zdroji na obrázky v druhém svazku, úplně poslední strana, zahrledne jméno Katka Němčková, její fotka je na str. 34 vpravo. Byla to při čtení podobná radost, jako zaslechnout českému v seriálu Stargate Atlantis (spinoř slavného seriálu Stargate CG). Nesním opomenout zmínku na Toma Stopparda, str. 92, rodáka ze Zlína, byt rodina emigrovala do Anglie v jeho útlém věku. Ještě jsem si poznačil jedno jméno, moment, ano, samozřejmě, Mendelovy zákony na str. 215 prvního svazku. Snad jsem nikoho dalšího neopomenul.

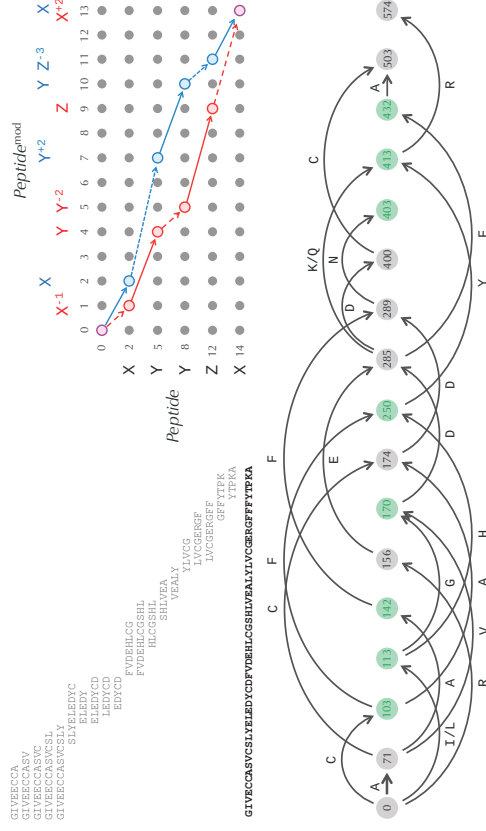
Poznámka závěrem. U této knihy, ale i obecně, jsem si říkal, že číslo strany by bylo lepší mít v záhlaví než v zápatí. Když totiž rychle procházíte PDF a máte náhled na stranu v zobrazení větším než je obrazovka, tak vidíte záhlaví, zápatí nikoliv. Možná tip pro bulletin? Tam to máme stejně. Byť to je jistý standard (ještě z *historických dob*) tištěných knih, ale tam je de facto jedno, jestli se díváte při rychlém listování na horní či spodní roh stránky.

Hezké bádání a suý v {A,C,G,T}! Poslední sazební ukázka je ze str. 185 prvního svazku.



Kapitola 11: Was *T. rex* Just a Big Chicken?

Computational Proteomics. Kapitola uvozuje práci s identifikací a výpočty kolem spekter peptidů. Přechíst kapitolu není těžké, ale naprogramovat výzvy a neztratit se v nich, to už těžké je. Na závěr ze světa sazby jsem si pro vás opět vybral jednu textovou ze str. 237, jedná se opět o jistý výsledek (seskládání peptidů u inzulinu, výsledkem je sekvence aminokyseliny, viz práce Fredericka Sangera) a dvě grafické ze str. 270 (dvě modifikace jednoho peptidu) a 243 (DAG; orientovaný acyklický graf).



Závěr každé kapitoly je jisté slovní shrnutí zdrojů a vývoj v dané oblasti.

Pokud jste víc matematici a statistici než „bio“, kniha nabízí nespočet úloh, zvlášť u kapitol 10 a 11 se protáčí panenky. Pokud jste víc programátoři, tak vězte, že biologické úlohy jsou převedeny na zadání zpracování textu, a to už bude vaše doména. Pokud nejste (ještě) ani jedno ze zmíněných, zkuste si pár úloh jen tak, pro zábavu z řešení stylu „Heuréka, ono mně to vyšlo!“ či si knihami prolustujte, pro inspiraci z hezky vysazené knihy.

Kniha tise nabízí možnost si úlohy vyzkoušet v *libovolném* programovacím jazyce. To ale nevylučuje užití *hotových* programů či knihoven, např. v R, resp. s pomocí BioConductoru. Věřím tomu a ověřil jsem si to, že na mnoha místech to jde, ale jádro knihy k tomu nesměřuje. Myslím si však, že člověk nemusí být zrovna druhý Donald E. Knuth (TAOCP), John Carmack (Doom, Wolfenstein 3D), Linus Torvald (Linux, Git), Dennis Ritchie (C), Steven Halim (knihla Competitive Programming), či Gennady Korotkevich v současnosti

Tabulka 1: Hodnoty původní a fuzifikované časové řady.

i	t_i	x_i	x_{il}	x_{im}	x_{ir}
1	1	1,7	0,75	1,7	2,00
2	2	1,7	1,25	1,7	2,00
3	3	2,0	1,50	2,0	2,50
4	4	2,4	1,75	2,4	2,75

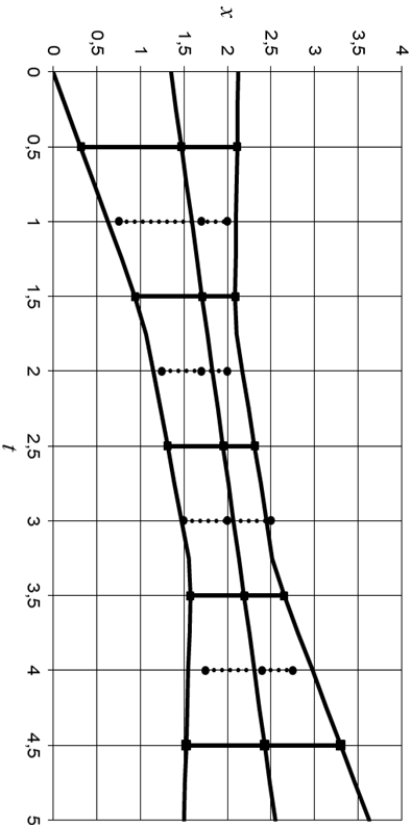
Tabulka 2: Hodnoty fuzzy časové řady a jejich odhady v čase t .

i	t	x_{il}	x_{im}	x_{ir}	$\hat{x}_{il}$	$\hat{x}_{im}$	$\hat{x}_{ir}$
1	1	0,75	1,7	2,00	0,6250	1,59	2,1000
2	2	1,25	1,7	2,00	1,1500	1,83	2,1750
3	3	1,50	2,0	2,50	1,4750	2,07	2,4500
4	4	1,75	2,4	2,75	1,5500	2,31	2,9750
---	0,5	---	---	---	0,3125	1,47	2,1125
---	1,5	---	---	---	0,9375	1,71	2,0875
---	2,5	---	---	---	1,3125	1,95	2,3125
---	3,5	---	---	---	1,5750	2,19	2,6500
---	4,5	---	---	---	1,5250	2,43	3,3000

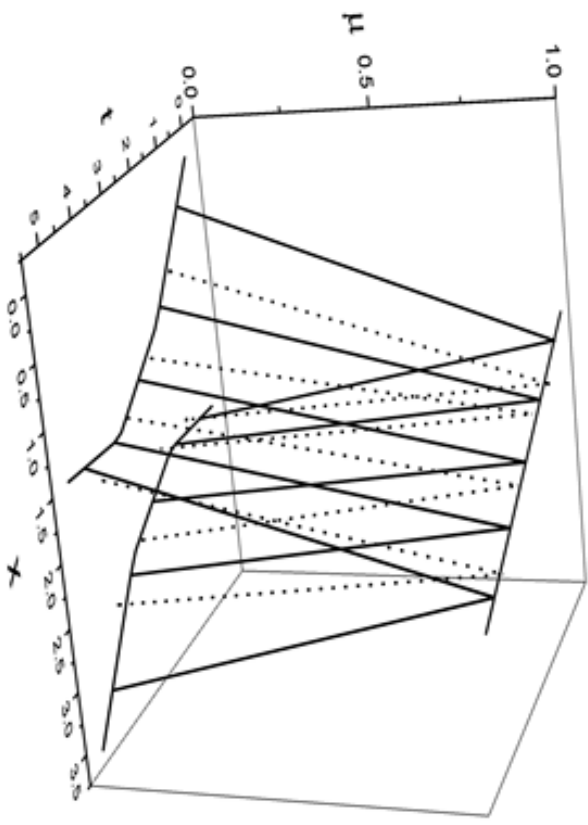
- vybraných nosičů a hlavních hodnot odhadů (svíslé plné čáry) z Tabulky 2 pro $t = 0,5; 1,5; 2,5; 3,5; 4,5$,
je na Obrázku 2.

Na Obrázku 3 je 3D graf odhadu fuzzy regresní přímky (trendu), zadané fuzzy časové řady a odhadů fuzzy hodnot řady časové řady vzhledem k funkci příslušnosti.

Ze získaných výsledků je zřejmé, že odhady fuzzy regresních koeficientů a odhad (vyrovnaní, aproximace) fuzzy hodnot fuzzy časové řady respektují nepřesné hodnoty pozorované časové řady. V daném příkladu ale nosič odhadu fuzzy regresního koeficientu β_2 obsahuje záporné i kladné hodnoty, takže daná časová řada může mít klesající anebo naopak rostoucí trend. Odhady pozorovaných fuzzy hodnot (svíslé úsečky) časové řady dané pásmem mezi dolní a horní lomenou čarou na Obrázku 2 i na Obrázku 3 totiž obsahují všechny možné přímky vviadřující trend této řady.



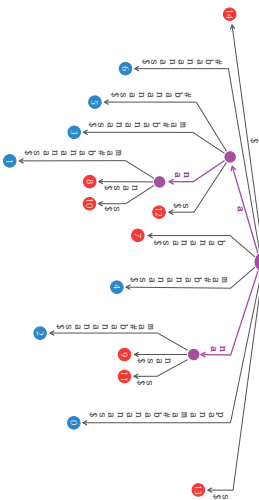
Obrázek 2: Graf nosičů fuzzy časové řady a odhadů hodnot fuzzy časové řady.



Obrázek 3: 3D graf funkce příslušnosti hodnot fuzzy časové řady (tečkovaná čára) a odhadů hodnot fuzzy časové řady (plná čára).

Kapitola 9: How Do We Locate Disease-Causing Mutations?

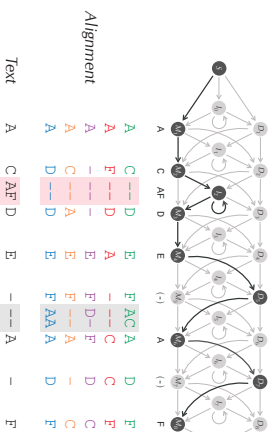
Combinatorial Pattern Matching. Kapitola uvádí do datových struktur a jejich rychlé prohlédávání: trie, suffix trie, suffix array. Velká část kapitoly je věnována transformaci Burrows-Wheeler a její inverzi. Pokud si to již někdo zkusil naprogramovat, tuší, že už jde v knize do tuhého. Na ukázkou ležké sazby jsem vybral SuffixTree ze str. 168.



Kapitola 10: Why Have Biologists Still Not Developed an HIV Vaccine?

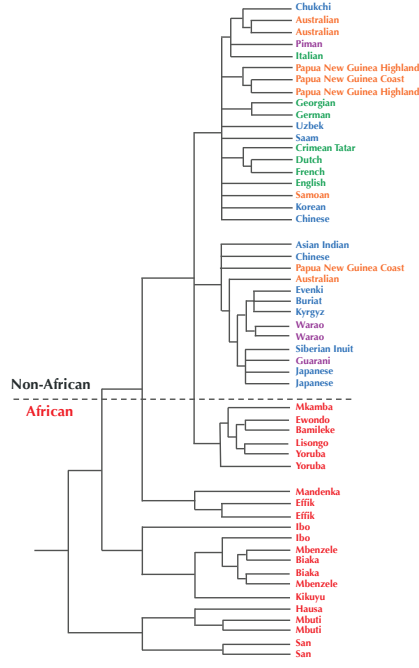
Hidden Markov Models. Skryté Markovské řetězce v plné síle plus Baum-Welch Learning. Za sazbu jsem nevěděl, co vybrat dítv, jestli barevné vzorce či reprezentaci HMM. Vybral jsem jednu textovou ukázkou (změněn v čase u proteinu gp120 u pacienta s HIV) ze str. 180 a jednu grafickou HMM ze str. 207 s volenou nepravděpodobnější cestou k danému problému.

VKILGHOER-NKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFN-----NSTES-----DTITTL
VKILGHOER-NKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFN-----NSTDNG-----DTITTL
VKILGHOER-NKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFN-----NSTESNN-----DTITTL
VKILGHOER-NKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFN-----TGNGESYNGQENGITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTGNS--TNIT--G.DG--NDITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTG--NSTE--G.HG--DDITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTG--NSTE--G.HG--DDITTL
VKILGHOEG-NKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFN--IR-----NSTESNNGQNDITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTG--NSTE--G.DG--NDITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTG--NSTE--G.DG--NDITTL
VKILRHOEGKNKTIIFNCPSSGGDEIVMHSNCGSEFFYCNTIQLEFNNTG--NSTE--G.DG--NDITTL



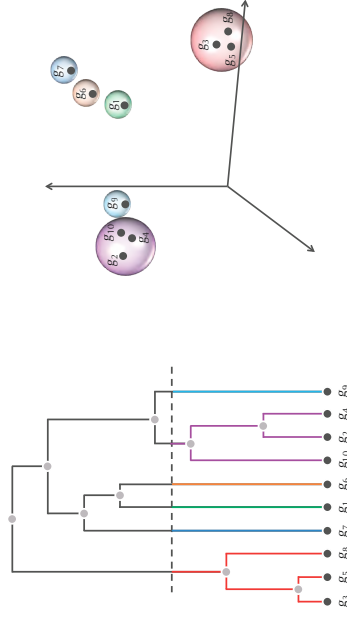
Kapitola 7: Which Animal Gave Us SARS?

Evolutionary tree reconstruction. Kapitola začíná druhý svazek a uvádí do evolučních stromů a jejich konstrukci. Podobnou konstrukci známe z ro-
dokmání, jen je to ve větším měřítku a vztahy se teprve hledají. Za sazbu
isem vybral evoluční strom ze str. 64, jen otočený skrz úsporu místa.



Kapitola 8: How Did Yeast Become a Wine Maker?

Clustering Algorithms. Shlukování je náplní této kapitoly. Mezi algoritmy najdeme starý dobrý známý Lloydův pro k -means a CAST. Poté se kapitola zabývá hierarchických shlukováním. Za sazbu jsem vybral strom ze str. 105 z 3D reprezentací. Na str. 104 a 107 upoutají tabulky se symetrickými hodnotami. Někdy se hodnoty vynechávají, zde se autoři rozhodli pro odlišnou (šedou) barvu.



6. Závěr

Slova trend, aproximace a predikce jsou v ekonomických úvahách a modelech velmi často používané a z hodnot vyjadřujících jejich velikost se vyvozují různé víceméně hlubokomyšlné závěry. Jak však ukazuje uvedený, byť jenom školní příklad, mohou být tyto závěry zcela falešné. Jde sice pouze o jednoduchý příklad, ale v praxi bývá situace ještě horší. Chráli jsme proto ukázat, že zanedbání nepřesností (chyb, šumu apod.) statistického nebo expertního zjišťování hodnot ekonomického nebo finančního sledovaného ukazatele a podobně i fyzikální veličiny v čase, může vést k velmi zkreseným závěrům. Popsané fuzzy modelování časových řad ale respektuje i původní hodnoty časové řady, pokud je chceme považovat za přesné.

Poděkování: Těchto výsledků bylo dosaženo v rámci specifického výzkumu FSI-S-20-6187 „Moderní metody aplikované matematiky“ Fakulty strojního inženýrství Vysokého učení technického v Brně. Příspěvek byl dále podpořen institucionálním financováním výzkumu Fakulty vojenského leadershipu Univerzity obrany v Brně.

Literatura

- [1] Anděl, J. (2011): *Základy matematické statistiky*. 3. vyd. Praha: Matfyzpress, 2011. 358 s. ISBN 978-80-7378-162-0. *cit. 4*
- [2] Karpíšek, Z., Lacinová, V., Sadovský, Z., Schneider, A. (2016): Is the Increasing Trend Always Really Increasing? MENDEL 2016 – 22th International Conference on Soft Computing. Brno, 2016. *Mendel Series*, Volume 2016, p. 229–234. ISSN 1803-3814, ISBN 978-80-214-5365-4. *cit. 4*
- [3] Karpíšek, Z., Dražanová, M., Lacinová, V. (2019): Lineární regresní model intervalové časové řady. *ACTA STING*, 1/2019, p. 24–35. ISSN 1805-6873. *cit. 4, 7*
- [4] Klir, G. J., Yuan Bo. (1995): *Fuzzy Sets and Fuzzy Logic – Theory and Applications*. New Jersey: Prentice Hall, 1995. ISBN 0-13-101171-5. *cit. 6*
- [5] Mareš, M. (1994): *Computation over Fuzzy Quantities*. Boca Raton: CRC Press, 1994. ISBN 978-08-493-7635-1. *cit. 6*
- [6] Montgomery, D. C., Runger, G. (2010): *Applied Statistics and Probability for Engineers*. 5th ed. New York: John Wiley, 2010. 784 s. ISBN 978-0-470-05304-1. *cit. 4*
- [7] Tsay, R. S. (2010): *Analysis of Financial Time Series*. New York: John Wiley, 2010. 672 s. ISBN 978-0-470-64455-3. *cit. 4*

RECENZE KNIHY: COMBINATORICS – ANCIENT & MODERN
BOOK REVIEW: COMBINATORICS – ANCIENT & MODERN

Pavel Striz

E-mail: pavel@striz.cz

Robin Wilson, John J. Watkins (editors): *Combinatorics – Ancient & Modern*, 1st edition, Oxford University Press, United Kingdom, x+381 pp., 2013. ISBN 978-0-19-965659-2.

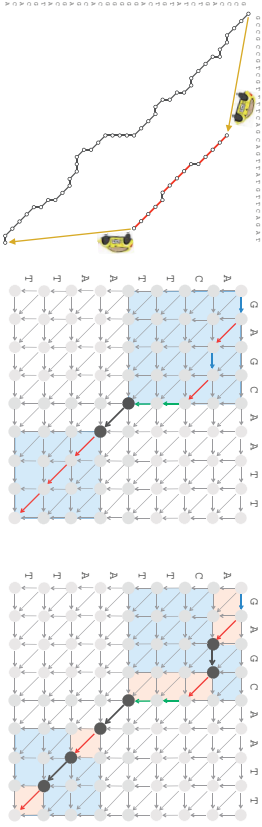
COMBINATORICS: ANCIENT & MODERN	
EDITED BY ROBIN WILSON & JOHN J. WATKINS OXFORD	
Part I Introduction Two thousand years of combinatorics DONALD E. KNUTH	9. Partitions GEORGE E. ANDREWS
Part II Ancient Combinatorics	10. Block designs NORMAN BIGGS AND ROBIN WILSON
1. Indian combinatorics TAKANOBU KUSUDA AND KIM PLOKKER	11. Latin squares LAAS DOWLING ANDERSEN
2. China ANDREA BRÉARD	12. Enumeration (18th–20th centuries) E. KEITH LLOYD
3. Islamic combinatorics AMIR DIBBAR	13. Combinatorial set theory IAN ANDERSON
4. Jewish combinatorics VICTOR J. KATZ	14. Modern graph theory LOWELL BEINEKE AND ROBIN WILSON
5. Renaissance combinatorics ERHARDB KNOBLOCH	Part IV Aftermath A personal view of combinatorics PETER J. CAMERON
6. The origins of modern combinatorics ERHARDB KNOBLOCH	Note on contributors Picture sources and acknowledgments Index
7. The arithmetical triangle A. W. F. EDWARDS	
Part III Modern Combinatorics ROBIN WILSON	

Dobrovolně se přiznávám, že jsem si nebyl jistý, jestli knihu číst. Zajímá mě kombinatorika, zvláště ta rekreační, ale historie už mň šálek kafe není. Když jsem jako student prvního ročníku na vysoké škole začal pomáhat se skripty do Metod statistické analýzy (někde nazýváno jako Statistika 1, I či A), Lada Rytíř trval na tom, že tam úvodní kapitola s historií být musí, že ta je důležitá, a páli na mě jedno jméno za druhým, že ti tam rozhodně být musí. Že na tom kategoriálně trvá. O každém hovořil, jako kdyby s danou osobou včera večer seděl na konferenci u piva. Celý Ladík.

A tak dobře, říkáš si, zkusím to. Začnu číst a uvidím. Knuth knihu uvádí (přetisk z TAOCP – Dva tisíce let kombinatoriky), asi mě chytl a nepustil. Pak jsou následující dvě části (Ancient Combinatorics; Modern Combinatorics) rozpracovány; prakticky na stranu každá polovina knihy. V druhé části: indická, čínská, islámská, židovská a renesanční kombinatorika plus počátky

Kapitola 5: How Do We Compare Biological Sequences?

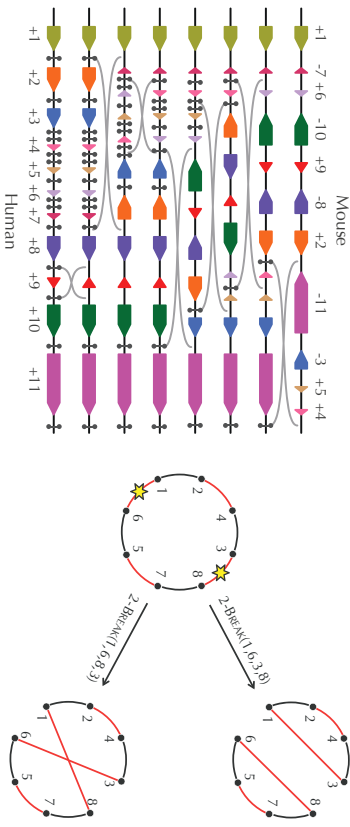
Dynamic Programming. Mé oblíbené úlohy, jak s minimem změn z jednoho řetězce dostat druhý. Tato úloha mě dovedla k této knize. Uvede dále i do variant lokální a globální a úlohy s více než dvěma řetězci. Obrázky z teorie grafů jsou čtenáři jistě dobře známé, za sazbu jsem vybral obrázek ze str. 260, 276 a za sazbu textu výsledek zarovnání tří řetězců a jejich barevné odlišení ze str. 290. Na vysvětlenou. Červená značí shodu v třech případech, zelená, modrá a hnědá v příslušných dvou případech, u černé není shoda a divís je vložení prázdného znaku do původního řetězce.



VAFD,IGTCMEFPV,LLGGEE,HLIVQKE,ITYA,DEIAHY,IKENH,IT,IK,IT,TP,SL,EFTT,IVN,TA, -AFDVSAGDEFA,RA,LLT,GG,QL,IVCNE,IK,MDP,DP,PA,SL,IV,IA,IK,MD,IT,TF,FEAT,PA,LLV,IP,ME,Y,I, -IAFD,ASWEI,YA,P,LLN,GG,IT,GCIDY,TT,TD,IK,RA,EW,RO,HH,I,RG,AM,PP,AL,IK,OC,IV,SA, SPA,PA,NF,ES,RL,IV,LV,GG,EK,IT,IP,DV,IA,R,KM,G,ENE,-F,IN,H,Y,G,T,EA,T,IG,A, -YEÖK,LD,IS,ÖQ,LL,IV,SD,SC,ME,DE,KT,IV,SR,FG,ST,IR,IV,NS,Y,G,V,T,EA,C,IDS, ---PTM,ISS,LE,IL,IF,AA,GD,R,LS,ÖD,ALA,RA,AV,GS,V,-Y,-NA,Y,G,P,T,EN,T,V,LS

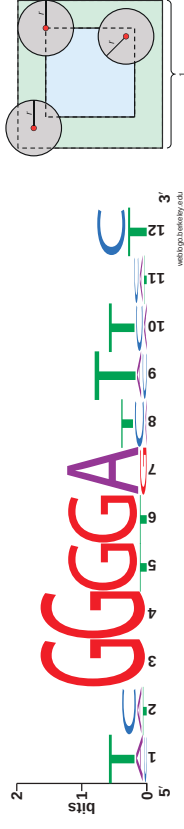
Kapitola 6: Are There Fragile Regions in the Human Genome?

Combinatorial Algorithms. Úvod do kombinatoriky, přesněji permutací, jak získat chromozom z jiného. Za svět sazby jsem vybral znázornění, jak z myši udělat člověka ze str. 302 a transformaci genomu ze str. 338.



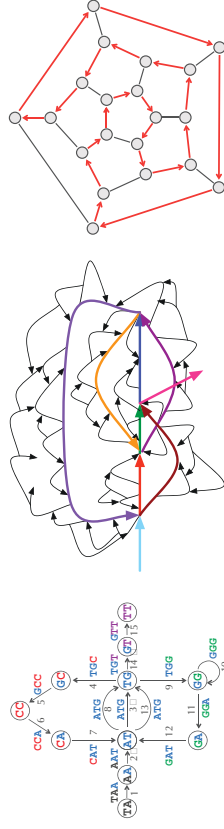
Kapitola 2: Which DNA Patterns Play the Role of Molecular Clocks?

Randomized Algorithms. Hledání vzorů a jejich ohodnocení, Hamming Distance, Gibbs Sampling. Za sazbu jsem vybral obrázek ze str. 74 a pravý obrázek ze str. 109.



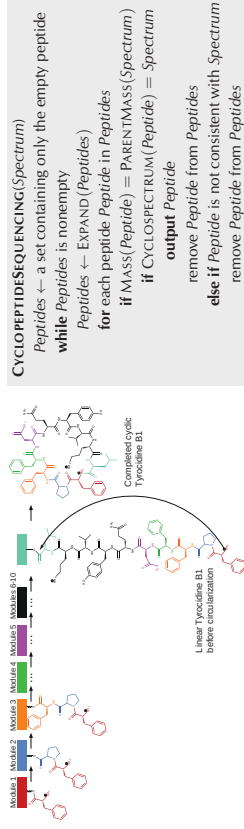
Kapitola 3: How Do We Assemble Genomes?

Graph Algorithms. Úvod do teorie grafů, rekonstrukce řetězce, Hamiltonian Path, De Bruijn Graph. S úsměvem lze říci, že se jedná o slavné FBI případy, tedy seskládání důležitých dokumentů prohnanych skartovačkou. Zkusení hackerů ví, že dokumenty je potřeba pálit, nikoliv skartovat. U této kapitoly si člověk připadal jako správný detektiv. Za sazbu jsem vybral obrázky ze str. 134, 161 a 175.



Kapitola 4: How Do We Sequence Antibiotics?

Brute Force Algorithms. Kapitola uvádí sekvencování peptidů, konvoluci. Za sazbu jsem vybral vysazení proteinů ze str. 189 a zdrojového kódu ze str. 196.



moderní kombinatoriky, v třetí části pak dle kombinatorického užití: počátky teorie grafů, celočíselný rozklad, design experimentů, latinské čtverce, výčty, kombinatorická teorie množin a současná teorie grafů. V poslední části, řečnické Doslov (Aftermath), Peter J. Cameron píše o svém osobním pohledu na kombinatoriku, hlavně její vývoj.

Knihu uzavírají medailonky autorů, zdroje obrázků, poděkování a společný rejstřík termínů a osob.

Obrázky z původních knih začínají jednotlivé kapitoly a častokrát doprovází text. Ačkoliv se na knize podílela celá řada autorů, během čtení to nepoznáte. Jazyk zní, jako kdyby to psal jeden autor. Poklona editorům.

Jmenovitě tedy: Donald E. Knuth (Úvod), Takanori Kusuba (1), Kim Plofker (1), Andrea Bréard (2), Ahmed Djebbar (3), Victor J. Katz (4), Eberhard Knobloch (5, 6), A.W.E. Edwards (7), Robin Wilson (8, 10, 14 plus editor), George E. Andrews (9), Norman Biggs (10), Lars Døvling Andersen, E. Keith Lloyd (12), Ian Anderson (13), Lowell Beineke (14), Peter J. Cameron (Doslov) a John J. Watkins (jen jako editor).

Kniha je poutavá z toho pohledu, která osoba a co neudělala správně, nebo udělala správně a kam až se dostala, a byla-li či nebyla-li dále prohlubována, a po jaké době. Kniha je ojedinelá a první svého druhu. V Předmluvě editorů hovoří o tom, že kniha má dva účely: jako první velký průzkum historie kombinatoriky tohoto rozsahu a v jedné knize má být shrnuta rešerše zdrojů, které jinak nejsou dostupné široké veřejnosti.

Nášinec zajásá, protože na str. 348 je zmínka o Otakaru Borůvkovi, v literatuře jako 9. zdroj, viz str. 350. A nejen to, na té samé straně u 4. zdroje máme Václava Chvátala. Plus malá poznámka na str. 369 o Mendelovi.

Osobně mě potěšila zmínka v kapitole 11 na str. 279 a 280 o sudoku a výpočtu různého počtu řešení sudoku bez a se zahrnutím vybraných elementárních transformací. Řeším jejich ranking-unranking kombinatorickou úlohu.

(Nejen) pomocí kombinatoriky bylo v březnu (The Hat, zahrnutí překlopení dílku) a v květnu 2023 (The Spectre, už bez překlápění) završeno hledání aperiodického dílku, který zaplní plochu bez opakování vzoru. Byl to dlouhodobý otevřený problém. Doufám, že tuto informaci uvidím v dalším vydání této knihy. Těším se!

Recenzi knihy zakončím slovy z Úvodního slova Ronalda Grahama:

One of the most compelling instincts that human beings have is the irresistible urge to look for patterns: this is apparent from the earliest attempts of our ancestors to understand the world around them. Mathematics has often been described as the science of patterns, and perhaps more than any other mathematical field, this represents the heart and soul of combinatorics.

ROZŠÍŘENÁ RECENZE KNIHY: BIOINFORMATICS ALGORITHMS – AN ACTIVE LEARNING APPROACH, VOL. I, II

EXTENDED BOOK REVIEW: BIOINFORMATICS ALGORITHMS – AN ACTIVE LEARNING APPROACH, VOL. I, II

Pavel Striž

E-mail: pavel@striz.cz

Phillip Compeau, Pavel Pevzner: *Bioinformatics Algorithms – An Active Learning Approach*, 2nd Edition, Active Learning Publishers, USA, Volume I (xxv+355 pp.), II (xxiii+291 pp.), 2015. Vedle titulních stran svazků jsou ilustrace uvozující kapitoly 2 a 10.



Motto: *Find the {most likely, mass, minimum, maximum, length, longest, shortest, highest-scoring, reverse, pattern, optimal, position, center, frequent, circular, nearest,...} in a string.*

Před lety jsem hledal server, který by nabízel sérii úloh na procvičování programování, ale aby to byl nějaký rozumný počet a uzavřený. Plus abych mohl vydat něco typu *Sbírka řešených úloh* s odkazy na zdroje. Dlouhodobě se zabývám zpracováním textu a na konferenci OSSConf v Žitné jsem zaslal: „... my na to v textu na webe používáme maticí vzdáleností a funguje to naozaj dobře...“. Říkal jsem si, že to už jsem někde viděl či čel, že se na to musím doma podívat. Edit distance od Levenshteina z 1966 „bylo naozaj to, o čem bratia hovorili“. Hledal jsem nějaké příklady a ukázky a za chvíli jsem byl na webu <http://rosalind.info> a viděl i různé modifikace tohoto problému, viz sekce *Bioinformatics Textbook Track*. Úlohy mají na webu a není omezení na programovací jazyk ani čas (jak dost často bývá zvykem u soutěžního programování) a mají tam úvod do Pythonu i algoritmizace.

Začal jsem úlohy postupně řešit, pak jsem se zasekl. Tak jsem dohledal jejich knihy a znovu řešení úloh obnovil. V době psaní recenze mi ještě pár tuctů úloh chybí, ale ty už jsou obtížností z jiného světa, ale je vidět světlo na konci tunelu. A hlavně jsem na druhý pokus (první asi zapadl mezi emailové spamy) dostal souhlas autorů na takový knižní počín.

O co tedy v knižkách jde? Jedná se de facto o skriptu na bioinformatiku podpořené serverem na úlohy z knihy plus celý blok úloh nezmíněné v knize, ale též z bioinformatiky. Obrovskou výhodou je, že po uznání úlohy jako vyřešené je k dispozici diskuzní fórum k dané úloze, dá se častokrát narazit na zajímavé myšlenky a postupy plus řešení v nejprůzračnějších jazycích.

Dvoustavková kniha mi za ty roky přirostla k srdci. Věděl jsem, že závažné problémy budou těžké, dokonce Pavel Pevzner o tom někde psal, že to tak dělá u svých studentů, že přes těžké (i některé nevyřešené) problémy se z nich snaží dostat to nejlepší, říkáme jim *Code Challenges*. Snad se v budoucnu se všemi vyřešenými úlohami pochlubím.

Za zmínku stojí i jejich YouTube kanál, kde řeší vybrané úlohy. Odkazy a prezentace, viz <http://bioinformaticsalgorithms.org>. V knize vás potěší *Detours*, tedy čtení rozšiřující a doplňkové. Nápořky k úlohám lze hledat pod *Charging Stations*. Kniha vyzývá i k řešení reálných problémů, viz *Final Challenges*.

Pokud se zabýváte alespoň trochu typografií a sazbu, kniha pro vás nebude zklamáním ani z tohoto pohledu. V metadatach PDF souborů lze vyčíst, že je sázen TeXem. Obrázky (grafy, stromy, diagramy ap.) jsou vysázené ve větší případě v barvě a v prvotřídní kvalitě. Ve vzorcích zahledíte barevné výrazy, to byla jedna dobu má příjemná úloha automatizace pro LuaTeX. U jednotlivých kapitol příkladem vzorek s otázkou, jak to vysázet?

Kniha je rozdělena do 6+5 kapitol. Zkusím je stručně představit, jaké oblasti řeší a vždy několik problémů, které se v kapitole objeví.

Kapitola 1: Where in the Genome Does DNA Replication Begin?

Algorithmic Warmup. Úvod do četností, měření odlišností, hledání vzorů a práce s textovými řetězci skládající se ze znaků {A,C,G,T}. Musíme mít na paměti, že je to příjemná konvence pro čtenáře užívat znaky (8 bitů), efektivnější samozřejmě je pracovat se dvěma bity, které nám čtyři znaky dokážou zakódovat. Na ukázkou sazby jsem vybral horní obrázek ze str. 21.

