Informační Bulletin

České Statistické Společnosti číslo 4, ročník 17, prosinec 2006

SBOHEM, PANE PROFESORE ...

Stanislav Komenda

Adresa: Univerzita Palackého v Olomouci

E-mail: komenda@upol.cz

Z výboru České Statistické Společnosti došla zpráva, že zemřel Ing. Josef Machek. Působil na katedře pravděpodobnosti a matematické statistiky matfyzu, Matematicko-fyzikální fakulty Univerzity Karlovy, od samotného založení fakulty v roce 1952. Byl tehdy čerstvým absolventem ČVUT, její fakulty s poněkud netradičním názvem – *Vysoká škola speciálních nauk*. Tu fakultu hned potom bolševici zrušili; studovaly se na ní obory *pojistná matematika, ekonometrie* a *matematická statistika*. Ty první dva byly vhozeny rychlokvašenými marxistickými učenci do pytle *buržoazních pavěd*, spolu s kybernetikou. Matematická statistika byla ušetřena a stala se základem zmíněné katedry na matfyzu.

Ing. Machek byl o sedm let starší než já. Setkali jsme se v roce 1954, kdy jsem byl, po jistých peripetiích, přijat ke studiu matematické statistiky. Naše kontakty podstatně vzrostly po třetím semestru, kdy pro nás skončila první část studia, společná s ostatními matematickými kombinacemi – a našeho vzdělávání se ujala katedra statistiky. Nebyla nijak rozsáhlá, a odborný asistent Ing. Josef Machek byl na ní nepřehlédnutelný. Přednášel a cvičil několik předmětů. Tím hlavním byla, myslím, teorie *statistického odhadování*.

Sešity s jeho výkladem jsem si schoval; přestože jsem se z podstatné části po promoci živil analýzou dat, moje pravděpodobnostní modely asociativního učení se bez metod odhadování neobešly. K přednáškám profesora Jaroslava Janko *Statistická kontrola jakosti* s námi pan asistent Machek konal cvičení. Asi jsme nebývali příliš soustředění – na páně profesorovu přednášku jsme totiž dobíhali udýchaní a propocení z tělocvičny *Maratón* sídlící v Křemencové ulici hned vedle hospody u Fleků. Tam ovšem bývalo dopoledne zavřeno; doplňovat tělesné tekutiny v tělocvičně vydané jsme museli v nedalekém Černém pivovaru na Karlově náměstí vedle rohové budovy techniky.

Odhaduji, že jsem prošel rukama několika desítek pedagogů. Jako žák a jako student. Podobně jako většina mých bližních. Paměťové záznamy téměř všech blednou a upadají v zapomnění. Zůstávají epizodky nepodstatného. Jenom několik málo učitelů si dokážu vybavit jako úplný obraz nebo dokonce jako postavu živou, se mnou hovořící, gestikulující, mnohorozměrnou, multidimenzionální. To poslední je navíc statistický odborný pojem. Ing. Josef Machek patří mezi ně. Byl to člověk mnoha talentů – rozhodně byl talentovaný matematik, schopný hluboké abstrakce. O tom svědčí i základní díla matematické statistiky jím přeložená. Měl i talent jazykový; pokud vím, komunikoval výborně anglicky; několik let přednášel matematickou statistiku na Kubě. Kromě zkušeností odborných i jazykových ze svého pobytu vytěžil i svou španělskou ženu. Nepřehlédnutelná byla jeho velkorysost, s níž rozdával své znalosti nám, talentem méně obdařeným. V prvních letech svého olomouckého působení jsem se občas jezdil za ním na katedru (sídlící ještě v ulici Ke Karlovu) radit. Patřil jsem tedy k těm jeho studentům, kteří ho okrádali o drahocenný čas i poté, co dostali v Karolinu svoje diplomy promovaných matematiků.

Ing. Josef Machek byl pro mne prototypem anglického gentlemana – nejenom pro své anglofilské záliby. Byl skromný tím nejopravdovějším způsobem; nikdy jsem ho nezaslechl, že by se svými znalostmi holedbal – a že bylo čím. Patřil navíc k několika málo lidem, které znám, kteří se vůbec nesnažili obložit svá jména akademickými tituly zepředu a zezadu. Zdá se, že si byl hluboce vědom, že hodnota člověka je v něčem jiném. Anglický byl i jeho smysl pro humor; opravdový, nenucený, neformální a přirozený. Měl ho vrostlý nejenom pod kůži. Jeho glosy se strefovaly *in medias res.* V jeho přítomnosti jsme se neváleli smíchy – vždycky jsme se však intelektuálně bavili – a zároveň se od něho učili. Ing. Josef Machek byl jedním z lidí, díky nimž byl matfyz v dobách komunistického absurdna oázou, v níž pramen rozumu nikdy nevyschl, protože pořád tady byli, kdo ho opatrovali a opatrovat učili. Byl jsem pyšný, když mi nabídl tykání. Bude se mi odcházet ze světa lehčeji, když vím, že se zase sejdeme.

REGULAČNÍ DIAGRAMY S ROZŠÍŘENÝMI MEZEMI

RNDr. Jiří Michálek, CSc.*, Ing. Josef Křepela**

* *Adresa:* Ústav teorie informace a automatizace AV ČR
Pod Vodárenskou věží 4, 182 07 Praha 8
E-mail: michalek@utia.cas.cz
** *Adresa:* Centrum pro jakost a spolehlivost výroby
Fakulta strojní ČVUT v Praze, Technická 4, 160 00 Praha 6
E-mail: krepela@atlas.cz

Abstrakt: V současné době se ukazuje, že celá řada výrobních procesů se nedá jednoduše regulovat pomocí klasických regulačních diagramů Shewhartova typu, a to především proto, že nelze dosáhnout konstantního průběhu parametru polohy a nebo díky malé úrovni inherentní variability procesu je možno připustit i pohyb v parametru polohy. Příspěvek rozvádí obecný přístup k řešení takové situace, která by zajistila navržení takových regulačních mezí, aby bylo možno dosáhnout předepsaného rizika výskytu falešných poplachů.

Klíčová slova: Shewhartovy regulační diagramy, riziko prvního druhu, normální rozdělení

1. Úvod

Pod vlivem praxe se ukazuje, že předpoklady na chování sledovaných měřitelných jakostních znaků vyžadované klasickým Shewhartovým přístupem jsou v reálném světě velice zřídka kdy beze zbytku splněny. Jedná se totiž o požadavky na normalitu rozdělení jakostního znaku, tedy typ rozdělení pravděpodobnosti $N(\mu, \sigma^2)$, jehož parametry z dlouhodobého pohledu se nemění v čase, tzn., že střední hodnota μ a úroveň rozptylu σ^2 jsou v čase konstantní. Tento požadavek se ukazuje být dosti nereálným, a proto se začaly sledovat i takové procesy, u nichž se může střední hodnota μ a rozptyl σ^2 měnit v čase a rovněž rozdělení pravděpodobnosti pro sledovaný znak nemusí být jenom normální. Pak ovšem pro sledování a řízení takového zobecněného procesu je nutné samozřejmě změnit i konstrukci regulačních mezí a výpočet koeficientů způsobilosti.

Zde se budeme zabývat následujícím modelem, který se zdá být velice vhodným pro použití v praxi. Nechť $\{x_{ij}\}_{j=1}^{n}$ je i-tá logická podskupina, n

je tedy její rozsah. Předpokládáme, že každé pozorování x_{ij} se dá aditivně rozložit na

$$x_{ij} = \mu_i + e_{ij},$$

kde μ_i představuje střední hodnotu i-té podskupiny a e_{ij} jsou vzájemně nezávislé náhodné chyby, které mají nějaké rozdělení pravděpodobnosti s nulovou střední hodnotou a s rozptylem σ^2, jeho úroveň se tedy může měnit v čase. Chování středních hodnot μ_i, $i = 1, 2, \ldots, k$ lze buď chápat jako systematickou chybu (např. μ_i lze popsat lineárním trendem či sinusovkou), nebo jako náhodné veličiny, které obecně tvoří nějakou časovou řadu, tzn., že nemusí být ani navzájem nezávislé. Lze ale předpokládat, že střední hodnoty μ_i, když je chápeme jako náhodné veličiny, budou nezávislé na chybách $\{e_{ij}\}$. Tyto chyby představují totiž vlastní, tedy inherentní variabilitu procesu.

Je zřejmé, že proces popisující chování jakostního znaku se de facto skládá ze dvou procesů $\{\mu_i\}$ a $\{e_{ij}\}$, které je nutno řídit každý zvlášť, neboť jsou nezávislé. Protože u inherentních chyb $\{e_{ij}\}$ předpokládáme nulovou střední hodnotu, není nutné se řízením tohoto parametru zabývat. Dostáváme se tedy namísto původních dvou regulačních diagramů dle Shewharta ke třem regulačním diagramům. První dva se týkají sledování aritmetických průměrů z podskupin, třetí sleduje úroveň inherentní variability. Pro sledování procesu zvolíme tedy následující statistiky:

$$\overline{x}_i = \mu_i + \frac{1}{n}\sum_{j=1}^{n} e_{ij},$$

kde budeme předpokládat, že $E\{\mu_i\} = \mu$ představuje dlouhodobou střední hodnotu procesu a

$$D\{\overline{x}_i\} = D\{\mu_i\} + \frac{1}{n}\sigma_i^2 = \tau_i^2 + \frac{1}{n}\sigma_i^2.$$

Pro sledování variability mezi veličinami $\{\mu_i\}$ se jeví rozumné klouzavé rozpětí

$$\delta_i = |\overline{x}_i - \overline{x}_{i-1}|,$$

které dává představu o chování variability procesu $\{\mu_i\}$. Pro sledování inherentní variability (tzn. uvnitř logických podskupin) vezmeme buď výběrovou směrodatnou odchylku

$$s_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \overline{x}_i)^2},$$

či výběrové rozpětí

$$R_i = \max_{1\le j\le n}\{x_{ij}\} - \min_{1\le j\le n}\{x_{ij}\}.$$

Nyní vyvstává problém, jak u takového procesu obecně stanovit regulační meze. Jestliže lze chování středních hodnot $\{\mu_i\}$ popsat jakožto náhodnou veličinu, pak obecně každé μ_i bude mít svoje rozdělení pravděpodobnosti dané hustotami $h_i(\cdot)$, kde

$$E\{\mu_i\} = \int_{-\infty}^{+\infty} x\, h_i(x)\,\mathrm{d}x = \mu, \qquad \int_{-\infty}^{+\infty} (x-\mu)^2\, h_i(x)\,\mathrm{d}x = \tau_i^2 = D(\mu_i)$$

a obdobně inherentní chyby e_{ij} budou pro každé i navzájem nezávislé a stejně rozdělené s hustotou $g_i(\cdot)$, kde

$$\int_{-\infty}^{+\infty} x\, g_i(x)\,\mathrm{d}x = 0, \qquad \int_{-\infty}^{+\infty} x^2\, g_i(x)\,\mathrm{d}x = \sigma^2.$$

Potom každá veličina x_{ij} pro pevné i má rozdělení pravděpodobnosti, které je konvolucí hustot $h_i(\cdot)$ a $g_i(\cdot)$, tedy dané hustotou

$$f_i(x) = h_i(x) * g_i(x) = \int_{-\infty}^{+\infty} h_i(x-z)\, g_i(z)\,\mathrm{d}z,$$

přičemž

$$E\{x_{ij}\} = \int_{-\infty}^{+\infty} x\, f_i(x)\,\mathrm{d}x = \mu, \qquad \int_{-\infty}^{+\infty} (x-\mu)^2\, f_i(x)\,\mathrm{d}x = \tau_i^2 + \sigma_i^2 = D(x_{ij}),$$

a dále podmíněná střední hodnota

$$E\,\{x_{ij}\,|\,\mu_i\} = \mu_i \quad \text{pro každé} \quad j = 1, 2, \ldots, n.$$

Toto vše plyne z předpokladu nezávislosti mezi $\{\mu_i\}$ a $\{e_{ij}\}$.

Pokud budou $\{\mu_i\}$ tvořit řadu ze vzájemně nezávislých veličin, pak $\{x_{ij}\}$ a $\{x_{i_0 j}\}$ pro $i \ne i_0$ jsou vzájemně nezávislé a pro pevné i jsou podmíněně nezávislé při podmínce μ_i.

Nadále budeme předpokládat nezávislost mezi veličinami μ_i a inherentními chybami $\{e_{ij}\}$. Tento předpoklad je v praxi velice často splněn. Hlavní otázkou je, jak takovýto proces popsaný výše uvedeným modelem sledovat pomocí regulačních mezí.

2. Konstrukce rozšířených mezí

Vyjdeme z toho, co požadujeme po regulačních mezích. Chceme, aby v případě statisticky zvládnutého procesu pouze předepsané procento bodů padlo mimo regulační meze. Obvykle se požaduje, aby toto procento, které vlastně představuje pravděpodobnost chyby l. druhu (tj. riziko, že budeme usuzovat na působení zvláštní příčiny variability, když žádná nebude existovat), bylo na úrovni 0,27 %. Takto je toto riziko stanoveno u klasických Shewhartových regulačních diagramů. To znamená, aby pro každou logickou podskupinu platilo, že aritmetický průměr $\overline{x}_i$ vypočtený z hodnot této podskupiny a rovněž tak vhodný odhad inherentní variability, např. výběrová směrodatná odchylka v rámci podskupiny rozsahu n

$$s_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\overline{x}_i)^2}$$

splňují následující: existují takové číselné hodnoty USL($\overline{x}$), LSL($\overline{x}$), USL(s), LSL(s), že pro ně platí

$$100\,P\{\mathrm{LSL}(\overline{x}) \le \overline{x}_i \le \mathrm{USL}(\overline{x})\} \ge 99{,}73\,\%$$
$$100\,P\{\mathrm{LSL}(s) \le x_i \le \mathrm{USL}(s)\} \ge 99{,}73\,\%,$$

a to pro každé i, tj. pro každou logickou podskupinu. Protože připouštíme ještě variabilitu mezi středními hodnotami $\{\mu_i\}$, tato je tedy „hlídána", jak již bylo řečeno výše; pomocí klouzavého rozpětí δ_i tak tedy rovněž i zde musí existovat meze USL(δ), LSL(δ) tak, aby

$$100\,P\{\mathrm{LSL}(\delta) \le \delta_i \le \mathrm{USL}(\delta)\} \ge 99{,}73\,\%$$

pro každé i. Pokud takovéto tři dvojice mezí lze nalézt, bude proces statisticky zvládnutelný a lze jej regulovat pomocí těchto mezí. Samozřejmě požadujeme, aby rozmezí mezi horní a dolní regulační mezí bylo co nejmenší. Toho lze dosáhnout následujícím postupem. Pro každou logickou podskupinu máme hustotu pravděpodobnosti $f_i(\cdot)$, která určuje vlastnosti pozorování v této podskupině. Pro každou statistiku, která slouží pro řízení procesu; tj. $(\overline{x}_i, \delta_i, s_i)$, je nutno spočítat příslušné rozdělení pravděpodobnosti a pak určit odpovídající kvantily pro pravděpodobnost 0,001 35 a 0,998 65. Tím např. pro $\overline{x}_i$ získáme hodnoty kvantilů $u_i(0{,}001\,35)$ a $u_i(0{,}998\,65)$. Regulační meze pro sledování $\overline{x}_i$ definujeme pak jako

$$\begin{aligned}\mathrm{LCL}(\overline{x}) &= \min_i u_i(0{,}001\,35)\\ \mathrm{UCL}(\overline{x}) &= \max_i u_i(0{,}998\,65).\end{aligned}$$

Zcela analogicky pro zbývající statistiky (δ_i, s_i). Při výpočtu hustoty rozdělení pravděpodobnosti pro $\overline{x}_i$ je nutno vycházet z toho, že veličiny $\{x_{ij}\}_{j=1}^{n}$ jsou podmíněně nezávislé při podmínce μ_i. To znamená, že podmíněná hustota pro $\sum_{j=1}^{n} x_{ij}$ při podmínce μ_i je rovna n-té konvoluci podmíněné hustoty $f_i(\cdot|\mu_i)$, tedy

$$f_i^{\sum x_i}(\lambda|\mu_i) = f_i^{[n]}(\lambda|\mu_i),$$

a tudíž

$$f_i^{\overline{x}_i}(\lambda|\mu_i) = n\, f_i^{[n]}(n\lambda|\mu_i).$$

Pak tedy nepodmíněná hustota pro $\overline{x_i}$ je rovna

$$f_i^{\overline{x_i}}(\lambda) = n \int_{-\infty}^{+\infty} f_i^{[n]}(n\lambda|z)\, h_i(z)\, \mathrm{d}z.$$

Pokud model zjednodušíme v tom, že na střední hodnoty $\{\mu_i\}$ nebudeme pohlížet jako na náhodné veličiny, ale na pevná čísla, i když obecně neznámá, pak pro každou logickou podskupinu lze spočítat klasické Shewhartovy regulační meze odvozené od kvantilů rozdělení pravděpodobnosti jednotlivých pozorování v rámci jedné logické podskupiny, kde tato hustota má tvar odvozený od modelu

$$x_{ij} = \mu_i + e_{ij},$$

kde ovšem $E\{x_{ij}\} = \mu_i$, $D\{x_{ij}\} = D\{e_{ij}\} = \sigma_i^2$. Globální či rozšířené regulační meze pro celý proces jsou pak odvozeny od těchto dílčích regulačních mezí opět jako výše uvažováním maxima pro UCL a minima pro LCL. Obdobné jsou vztahy i pro ostatní statistiky (δ_i, s_i).

Pokud na průměry podskupin $\{\mu_i\}$ budeme pohlížet jako na náhodné veličiny, pak za statisticky zvládnutý proces lze chápat takovou situaci, kdy lze chování $\{\mu_i\}$ popsat nějakým rozdělením, které se v čase „příliš“ nemění. Pak by bylo možno se dívat na $\{\mu_i\}$ jako na náhodný výběr z tohoto rozdělení a z informace obsažené v pozorováních $\{\mu_i\}$ by bylo možno získat odhady vhodných kvantilů pro stanovení regulačních mezí. Pokud by vhodné rozdělení bylo gaussovské $N(\mu, \sigma_n^2)$, pak nejlepšími odhady jsou aritmetický průměr z $\{\mu_i\}$ a

$$\hat{\sigma}_\mu = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(\mu_i - \hat{\mu})^2}$$

a pak

$$\mathrm{UCL}_\mu = \hat{\mu} + 3\hat{\sigma}_\mu, \quad \mathrm{LCL}_\mu = \hat{\mu} + 3\hat{\sigma}_\mu.$$

V případě nesymetrického rozdělení lze využít např. Johnsonovy křivky a použít postup pro odhad odpovídajících kvantilů založený na aproximaci pomocí Johnsonových křivek.

V tomto obecném případě samozřejmě vyvstává otázka, co budeme považovat za statisticky zvládnutý proces. Protože naším cílem je splnit přání zákazníka, které vždy vyjadřuje nějakou, pro něj přijatelnou, mez zmetkovitosti, je zřejmé, že statisticky zvládnutý bude takový proces, který je možno predikovat a zaručit tak zákazníkovi požadovanou úroveň kvality. Z tohoto požadavku ale jednoznačně plyne, že zvládnutý proces musí mít víceméně stacionární charakter, aby se dala úroveň kvality zaručit do budoucna. Nestacionární procesy lze špatně predikovat, neboť obvykle jejich pravděpodobnostní strukturu neznáme. Pokud tedy vyjdeme z našeho modelu, že střední hodnoty $\{\mu_i\}$ lze chápat jako náhodné veličiny, pak při jejich hustotách rozdělení pravděpodobnosti $\{h_i(\cdot)\}$ musí platit, že

$$\min_i u_i(0{,}001\,35) \geq u_0, \quad \max_i u_i(0{,}998\,65) \leq u_1,$$

neboli musí existovat takové, říkejme mu globální rozdělení pravděpodobnosti $g(\cdot)$, že jeho kvantily na úrovni 0,135 % a 99,865 % tuto roli plní. Toto rozdělení může pak sloužit jako typ rozdělení, které popisuje chování středních hodnot $\{\mu_i\}$. Tímto rozdělením může být např. i směs z jednotlivých rozdělení veličin $\{\mu_i\}$, jak bylo již uvažováno v případě s lineárním trendem, viz [3].

Tento přístup je v souladu s přístupem obsaženým v metodice „Six sigma“, kde chování středních hodnot logických podskupin je vymezen prostor o šířce 3σ, tzn. $l{,}5\sigma$ na každou stranu od cílové hodnoty, resp. od středu tolerančního pásma. Při srovnání těchto dvou přístupů to znamená, že $u_0 > T - 1{,}5\sigma$, $u_1 < T + 1{,}5\sigma$, kde T značí cílovou hodnotu. Pokud by se dalo chování středních hodnot popsat rozdělením $N(\mu, \tau^2)$, pak v optimálním případě by bylo $\mu = T$ a $6\tau \leq 3\sigma$, což dává vztah mezi variabilitou mezi podskupinami a inherentní variabilitou

$$2\tau \leq \sigma.$$

Pokud ale požadujeme přísnější požadavky na hlídání náhodného chování střední hodnoty podskupiny, pak bude samozřejmě i přísnější nerovnost mezi τ a σ. Např. při přísnějším požadavku, že

$$u_1 - u_0 \geq 8\tau,$$

pak rovněž musí být $8\tau \leq 3\sigma$ pro správně centrovaný proces. Když tedy lze popsat chování středních hodnot podskupin pomocí $N(\mu, \tau^2)$, pak klouzavé rozpětí $\delta_i = |\overline{x}_i - \overline{x}_{i-1}|$ bude mít odpovídající rozdělení pravděpodobnosti dané hustotou

$$h(x) = \frac{2}{\sqrt{\tau^2 + \frac{\sigma^2}{n}\sqrt{4\pi}}} \left[e^{\frac{x^2}{4\left(\tau^2 + \frac{\sigma^2}{n}\right)}} \right],$$

protože $E\{\overline{x}_i - \overline{x}_{i-1}\} = 0$, $D\{\overline{x_i} - \overline{x}_{i-1}\} = 2\left(\tau^2 + \frac{\sigma^2}{n}\right)$ a $\overline{x}_i$ a $\overline{x}_{i-1}$ jsou nezávislé s rozdělením $N\left(\mu, \tau^2 + \frac{\sigma^2}{n}\right)$. Jedná se o nejjednodušší případ tzv. překlopeného normálního rozdělení se střední hodnotou

$$E\{|\overline{x}_i - \overline{x}_{i-1}\} = \sqrt{\frac{\frac{4(\tau^2+\sigma^2}{n)}}{\pi}} = \frac{2}{\sqrt{\pi}}\sqrt{\frac{\tau^2 + \sigma^2}{n}}$$

a s rozptylem

$$\begin{aligned} D\{|\overline{x}_i - \overline{x}_{i-1}|\} &= 2\left(\tau^2 - \frac{\sigma^2}{n}\right) - \frac{4}{\pi}\left(\tau^2 - \frac{\sigma^2}{n}\right) \\ &= \left(\tau^2 - \frac{\sigma^2}{n}\right)\left(2 - \frac{4}{\pi}\right) \cong \left(\tau^2 + \frac{\sigma^2}{n}\right) \cdot 0{,}72675. \end{aligned}$$

3. Příklad k rozšířeným regulačním mezím

Na základě 42 podskupin po pěti pozorováních v každé podskupině byly vypočteny následující statistiky:

$\overline{\overline{x}} = 1\,021{,}467\,3$, $\hat{\sigma} = 0{,}002\,8$ uvnitř podskupin,
$\hat{\sigma}_{\mathrm{TOT}} = 0{,}004\,9$ celková výběrová směrodatná odchylka,
$\hat{\sigma}_{\overline{x}} = 0{,}004\,3$ výběrová směrodatná odchylka aritmetických průměrů podskupin.

Klasické Shewhartovy regulační meze počítané pouze na základě odhadu $\hat{\sigma}$ inherentní variability (tj. uvnitř podskupin) mají hodnoty

$$\mathrm{UCL} = 102{,}4710, \quad \mathrm{LCL} = 102{,}4636.$$

Protože tyto meze nelze použít pro regulaci, neboť zbytečně mnoho aritmetických průměrů $\overline{x}$ z podskupin padne mimo tyto meze, aniž by se s procesem něco dělo, je doporučeno pracovat s rozšířenými regulačními mezemi. Na základě čeho je vypočítat? Obecně je několik možností:

1. na základě $\hat{\sigma}_{\mathrm{TOT}}$ (namísto $\hat{\sigma}$ ve výpočtu klasických regulačních mezí);

2. na základě součtu rozptylů inherentní variability a variability mezi podskupinami, tedy
$$s = \sqrt{(\hat{\sigma})^2 + (\hat{\sigma}_{\overline{x}})^2};$$
3. na základě součtu směrodatných odchylek $\hat{\sigma} + \hat{\sigma}_{\overline{x}}$;
4. podle firmy Ford, kde se uvažuje posunutí regulačních mezí o jistou konstantu Δ, např. lze volit $\Delta/2 = 3\hat{\sigma}_{\overline{x}}$ či pevnou hodnotu bez ohledu na pozorování, blíže viz [4].

Vyjděme z modelu

$$\begin{aligned} x_{ij} = \mu_i + e_{ij} \quad & i = 1, 2, \ldots, k \quad \text{počet podskupin,} \\ & j = 1, 2, \ldots, n \quad \text{rozsah podskupiny.} \end{aligned}$$

Lze-li se na chování procesu dívat jako na realizaci náhodné veličiny s nějakou střední hodnotou a rozptylem, pak

$$\begin{aligned} E\{x_{ij}\} &= E\{\mu_i\} \qquad\qquad j = 1, 2, \ldots, n, \\ D\{x_{ij}\} &= D\{\mu_i\} + \sigma^2, \end{aligned}$$

kde σ^2 je úroveň inherentní variability uvnitř podskupin. Bude-li $D\{\mu_i\} = \rho^2$ nezávislé na i (což je vlastně předpoklad pro statistickou zvládnutelnost procesu), pak
$$D\{x_{ij}\} = \rho^2 + \sigma^2,$$
a dále pro
$$\overline{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} = \mu_i + \sum_{j=1}^{n} e_{ij}$$
platí, že:

$$\begin{aligned} E\{\overline{x}_i\} &= E\{\mu_i\} \\ D\{\overline{x}_i\} &= \rho^2 + \frac{1}{n} \sigma^2. \end{aligned}$$

Odtud směrodatná odchylka pro $\overline{x}_i$ je $\sqrt{\rho^2 + \frac{1}{n}\sigma^2}$. Pak regulační meze jsou

$$E\{\mu_i\} \pm 3\sqrt{\rho^2 + \frac{1}{n}\sigma^2},$$

když rozdělení μ_i bude normální (či pro praxi přibližně normální).

Pro statistickou zvládnutelnost procesu je nutné předpokládat, že

$$E\{\mu_i\} = \mu,$$

pak

$$\begin{aligned}
\mathrm{UCL}(\overline{x}_i) &= \mu + 3\sqrt{\rho^2 + \frac{\sigma^2}{n}} \\
\mathrm{LCL}(\overline{x}_i) &= \mu - 3\sqrt{\rho^2 + \frac{\sigma^2}{n}}.
\end{aligned}$$

Jak tedy tyto meze, a tím i parametry ρ^2 a σ^2, odhadnout? Je zřejmé, že μ je celkový průměr pro všechna x_{ij}, tedy

$$\hat{\mu} = \overline{\overline{x}} = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}.$$

Pokud $\{\overline{x}_i\}$ budou přibližně normální na základě působení centrální limitní věty, pak velice rozumným odhadem pro jejich směrodatnou odchylku je statistika

$$\hat{\sigma}_{\overline{x}} = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left(\overline{x}_i - \overline{\overline{x}}\right)^2},$$

což vlastně odhaduje právě variabilitu mezi podskupinami. Protože platí základní vztah z ANOVA

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \left(x_{ij} - \overline{\overline{x}}\right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(x_{ij} - \overline{x}_i\right)^2 + n \sum_{i=1}^{k} \left(\overline{x}_i - \overline{\overline{x}}\right)^2,$$

pak

$$\frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} \left(\overline{x}_i - \overline{\overline{x}}\right)^2 = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} \left(x_{ij} - \overline{x}\right)^2 + \frac{1}{k} \sum_{i=1}^{k} \left(\overline{x}_i - \overline{\overline{x}}\right)^2,$$

neboli

$$\left(\hat{\sigma}_{\mathrm{TOT}}\right)^2 = (\hat{\sigma})^2 + \left(\hat{\sigma}_{\underline{x}}\right)^2,$$

čili

$$\hat{\sigma}_{\overline{x}} = \sqrt{(\hat{\sigma}_{\mathrm{TOT}})^2 - (\hat{\sigma})^2}.$$

Tohoto faktu lze využít pro výpočet odhadu směrodatné odchylky pro $\overline{x}_i$. V našem příkladu máme, že

$$\hat{\sigma}_{\overline{x}} \cong 0{,}004\,22,$$

což se shoduje s dřívějším odhadem $\sigma_{\overline{x}} \cong 0{,}00429$. Pak tedy „správné“ regulační meze pro aritmetické průměry $\{\overline{x}_i\}$ jsou

$$\begin{aligned} \mathrm{LCL}(\overline{x}) &= 102{,}467\,3 - 3 \cdot \hat{\sigma}_{\overline{x}} = 102{,}454\,43, \\ \mathrm{UCL}(\overline{x}) &= 102{,}4673 - 3 \cdot \hat{\sigma}_{\overline{x}} = 102{,}480\,17. \end{aligned}$$

Na rozdíl od „rozšířených“ regulačních mezí odvozených od celkové směrodatné odchylky, které jsou v tomto případě

$$\begin{aligned} \mathrm{LCL}^*(\overline{x}) &= 102{,}460\,7, \\ \mathrm{UCL}^*(\overline{x}) &= 102{,}473\,8, \end{aligned}$$

a pro regulování procesu se zřejmě nehodí, neboť je příliš bodů mimo tyto meze, i když se s procesem nic neděje. Počítání regulačních mezí pro $\overline{x}_i$ na základě $\hat{\sigma}_{\mathrm{TOT}}$ není správné, protože ignoruje $\hat{\sigma}_{\mathrm{TOT}}$ rozdělení pozorování do podskupin a chování $\overline{x}_i$ posuzuje pak na základě celkové variability procesu. Jednotlivá pozorování x_{ij} nelze totiž považovat za výběr z jedné a téže populace, když je statisticky významný rozdíl mezi σ_{TOT} a σ-způsobilostním (uvnitř podskupin). Další podrobnosti ke konstrukci rozšířených regulačních mezí lze najít v odkazech [1], [2], [3] a [5].

Otevřenou otázkou ale zůstává problém, kdy proces řídit pomocí klasických a kdy pomocí rozšířených regulačních mezí. Některé odpovědi na tuto otázku lze najít v [6], kde je řešena otázka testování stability procesu.

4. Závěr

Otázka, jak regulovat takové procesy, u nichž nelze zajistit pevné parametry normálního rozdělení, které popisuje chování sledovaného jakostního znaku na výrobku, je velice aktuální, neboť tento problém nastává především u dlouhodobějšího regulování procesu. Je nutno konstatovat, že tento problém není doposud v literatuře zcela dořešen a většina softwarů určených pro SPC s touto možností vůbec nepočítá.

Literatura

[1] MICHÁLEK J.: Procesy s rozšířenými regulačními mezemi. Research Report No. 1986, říjen 2000; Akademie věd České republiky, Ústav teorie informace a automatizace.

[2] MICHÁLEK J.: Zobecněné Shewhartovy regulační diagramy. Research Report No. 2057, říjen 2002; Akademie věd České republiky, Ústav teorie informace a automatizace.

[3] MICHÁLEK J.: Koeficienty způsobilosti a výkonnosti v případech rozšířených regulačních mezí. Research Report No. 2009, leden 2001; Akademie věd České republiky, Ústav teorie informace a automatizace.

[4] FORD MOTOR COMPANY: Verification of SPC Software-Test Examples. Q-DAS 1995.

[5] KŘEPELA J., MICHÁLEK J., TONAR J.: Konstrukce regulačních diagramů v případech, kdy nelze aplikovat Shewhartovy regulační diagramy podle ČSN ISO 8258. [HTML dokument]. Česká společnost pro jakost, říjen 2004, který je dostupný na adrese
http://www.csq.cz/cz/statisticke_metody_clanky.asp

[6] KŘEPELA J., MICHÁLEK J.: Testování stability procesu, předpoklad aplikace Shewhartových regulačních diagramů. [HTML dokument]. Česká společnost pro jakost, únor 2005, který je dostupný na adrese
http://www.csq.cz/cz/statisticke_metody_clanky.asp

APLIKACE GENETICKÝCH ALGORITMŮ V PROCESU PLÁNOVÁNÍ VÝROBY

Roman Kasal, Pavel Stříž, Jozef Říha

Adresa: Univerzita Tomáše Bati ve Zlíně, Fakulta managementu a ekonomiky

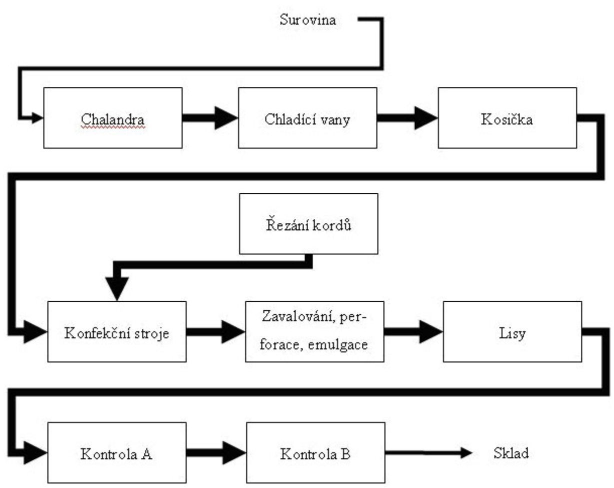
E-mail: `romankasal@centrum.cz, striz,riha@fame.utb.cz`

Abstrakt: Článek se zabývá problémem pocházejícím z reálného světa. Snaží se optimalizovat plán výroby ve firmě Mitas, a. s., s mnoha proměnnými. Článek popisuje optimalizační nástroj *Genetické algoritmy* v *Evolveru* a *Problém obchodního cestujícího* řešeného v programu *MATLAB*. Autoři věděli, že řešení nemusí být hledaným globálním minimem, ale i přesto nalezené řešení šetří 6/7 času klasického plánování výroby.

Klíčová slova: Plánování výroby, genetické algoritmy, Evolver, problém obchodního cestujícího, MATLAB.

1. Problém a popis výroby

Podúkol byl zpracovaný ve firmě Mitas, Zlín, jako součást rozsáhlejšího projektu zabývající se aplikací statistických metod v procesu plánování výroby motoplášťů v rozsahu 117 rozměrů. Konkrétně se jednalo o maximalizaci produkce při daných podmínkách. Nejprve bylo nutné detailně popsat výrobní halu, což vyústilo v základní vývojový diagram, viz Obrázek 1. *Vývojový diagram provozu na výrobu motoplášťů.* Analýza prozradila fakt, že množství produkce je závislé výhradně na výkonové neekvivalenci jednotlivých částí výrobní linky. Úzkým místem se ukázaly být lisy, tedy produkce byla závislá výhradně na využití 21 instalovaných lisů. Výkony a kapacity ostatních částí výrobní linky (včetně skladového omezení) nejsou rozhodující. Mají svůj systém optimalizace (ale relativně nezávislý) spočívající pouze v jednoduchém „ručním“ rozhodování úsekových nejvyšších pracovníků na základě požadavků výroby a odbytu. Problém se tedy týkal výhradně maximálního využití lisů, a to na jednom lisu s označením A, na kterém se vyrábí 9 rozměrů, pěti lisech B s 30 rozměry a 15 lisech C se 78 rozměry (různé skupiny rozměrů mají různé časy lisování). Lisy jsou v praxi využívány přibližně na 95 %. Cílem úkolu bylo maximální využití zmíněných lisů v rozmezí 14 dnů, po které je plánování uskutečňováno.



Obr. 1. Vývojový diagram provozu na výrobu motoplášťů.

2. Řešení

Pro řešení úkolu vzhledem k mnoha nezávisle proměnných byla zvolena metodika užívající genetické algoritmy.

Úkol byl řešen za pomocí softwaru *Evolver* od společnosti *Palisade*, který pracuje jako doplněk k *MS Excelu*. Řešení nebudeme příliš matematicky popisovat, protože chceme zdůraznit především praktickou aplikaci genetických algoritmů, protože nejsme ani matematici, ani statistici.

2.1. Účelová funkce a definice proměnných

Postup přípravy dat se odvíjel od účelové funkce, tj. funkce vyjadřující využití lisovacích strojů. Účelová funkce, kterou se snažíme maximalizovat, má pro jeden den tvar:

$Z = \sum_{i=1}^{117} x_i - \sum_{k=1}^{2} p_k w_k$, tzn., cílem je maximalizovat výstup provozní linky, což představuje výraz $\sum_{i=1}^{117} x_i$, a současně minimalizovat jednotlivé penalizace, tj. výraz $\sum_{k=1}^{2} p_k w_k$, kde p_k je určitá penalizace a w_k značí váhu, neboli důležitost penalizace v účelové funkci Z. Jinými slovy, cílem a výstupem je přesné rozplánování výroby, tedy přesné určení druhu a počtu výrobků (hodnoty x_i ve sloupci „T“ na Obrázku 2) na každý den a maximální využití časového fondu ($\sum_{k=1}^{2} p_k w_k$).

$$Z = \sum_{i=1}^{117} x_i - [(22{,}5 - A) + (22{,}5 \times 5 - B) + (22{,}5 \times 15 - C)]\, w_1 - \left[\sum_{i=1}^{117} v_i\right] w_2$$

$$T = A + B + C = \sum_{i=1}^{117} x_{iA} t_{iA} v_{iA} + \sum_{i=1}^{117} x_{iB} t_{iB} v_{iB} + \sum_{i=1}^{117} x_{iC} t_{iC} v_{iC}$$

představuje využitý časový fond, kde písmeny A, B, C jsou označeny lisy, viz popis problému, pro něž platí omezení například pro běžný pracovní den $A \in \langle 0; 22{,}5 \rangle$, $B \in \langle 0; 22{,}5 \times 5 \rangle$ a $C \in \langle 0; 22{,}5 \times 15 \rangle$ v hodinách. Protože den má 22,5 pracovních hodin a počet lisů A, B a C je 1, 5 a 15, pak omezení je právě násobkem počtu denních odpracovaných hodin. Další proměnné:

x_{iA} – počet kusů motopláště i, který je potřeba vyrobit podle programu *Evolver*;

t_{iA} – čas lisování motopláště i v hodinách;

v_{iA} – binární hodnota, která nabývá hodnoty „1“, pokud je třeba daný motoplášť vyrobit, a to jak podle potřeby vyrobit daný produkt na základě požadavku, tak podle programu *Evolver*;

w_1 – váha stanovená řešitelem, která určuje, jak je pro celkovou účelovou funkci důležité využití pracovní doby;

w_2 – váha stanovená řešitelem, která určuje, jak je pro celkovou účelovou funkci důležitý počet změn, tedy počet vyrobených rozměrů. Každá změna představuje prostoj, proto je dobré minimalizovat výraz $\left[\sum_{i=1}^{117} v_i\right]$.

2.2. Příprava dat v Microsoft Excelu

Postup přípravy dat se odvíjel od zvolení účelové funkce vyjadřující využití lisovacích strojů. Bylo tedy nutné ke každému ze 117 výrobku přiřadit čas lisování včetně manipulace, viz sloupec ‚C‘ v Obrázku 2. Další vstupní data byly počet jednotlivých lisů a jejich přiřazení k odpovídajícím rozměrům (sloupec ‚D‘), skladové zásoby (‚P‘) a denní požadavky (‚U‘) na množství pro jeden každý rozměr.

	A	C	D	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	počet lisů BOM 40"		2										
2	5	počet lisů BOM 30"							1	1				22,49
3	15	počet lisů BOM 40"							5	4				22,483
4									15	7	15	5	1	22,499
5														
6	č. výrobku	doba lisování + manipulace	počet lisů	sklad	GA výr.	GA ano/ne	nutná výroba	vyrobeno	potřeba	chybí	počet hodin na lis	počet hodin na lis	počet hodin na lis	celkem hodin
7	21212	15,3	15	50	1	0	0	0		0	0	0	0	0
8	22021	15,3	5	0	27	0	0	0		0	0	0	0	0
11	22041	15,3	5		335	0	1	335	320	0	0	42,71	0	42,713
12	22043	22,3	5		50	0	0	0		0	0	0	0	0
13	22049	22,3	5		56	0	0	0		0	0	0	0	0
14	22051	15,3	5		314	0	1	314	300	0	0	40,04	0	40,035
40	24432	32,3	15		16	0	0	0		0	0	0	0	0
41	24532	15,3	15		403	0	1	403	400	0	51,4	0	0	51,383
42	24544	15,3	15		66	0	0	0		0	0	0	0	0
44	24643	22,3	15		500	0	1	500	500	0	92,9	0	0	92,917
45	24649	32,3	15		17	0	0	0		0	0	0	0	0
46	24796	22,3	15		160	0	1	160	130	0	29,7	0	0	29,733
79	25141	17,3	1		68	0	0	0		0	0	0	0	0
80	25250	17,3	1		156	0	1	156	155	0	0	0	22,49	22,49
91	24087	22,3	15		27	0	0	0		0	0	0	0	0
92	24134	32,3	15		71	0	0	0		0	0	0	0	0
93	24167	22,3	15		301	0	1	301	300	0	55,9	0	0	55,936
94	24403	32,3	15		300	0	1	300	300	0	80,8	0	0	80,75
115	52008	15,3	5		30	0	0	0		0	0	0	0	0
116	52010	15,3	5		205	0	1	205	150	0	0	26,14	0	26,138
117	61011	15,3	5		55	0	0	0		0	0	0	0	0
121	52020	17,3	1		12	0	0	0		0	0	0	0	0

Obrázek 2. Výřez z Microsoft Excelu.

Princip řešení spočíval v určení maximálního počtu vyrobených kusů při využití časového fondu 22,5 hodin, tedy tří směn po 7,5 hodinách. Jelikož jde právě o řešení na základě úzkého místa, lisů, které byly skutečně využity na téměř 100 %, neuvažujeme další úseky provozu. Dále princip požaduje, aby bylo možné stanovit výrobu produktů, které je nutno bezpodmínečně vyrobit. To splňuje sloupec ‚U‘ („potřeba“). Programu *Evolver* byl nabídnut sloupec ‚Q‘ pro stanovení počtu vyrobených kusů zvoleného produktu. Sloupec ‚R‘ byl programu *Evolver* také nabídnut, aby binárně rozhodl o skutečné výrobě položky výroby z důvodu lepšího a rychlejšího nalezení řešení okamžitým zvolením takového výrobku, který vyplní kapacity volného lisu bez uvažování celých čísel ve sloupci ‚Q‘, což by mohlo vytvářet závislosti zpomalující účinnost algoritmu.

Protože sloupec ‚U‘, tedy sloupec, který je uživatelem ručně nastaven, určuje nutnost výroby daného výrobku (je zřejmě ten nejdůležitější), bylo potřebné přizpůsobit model, aby tuto skutečnost zohlednil a popřípadě zvýšil vhodnou váhou její důležitost. Řešením byl sloupec ‚S‘ („nutná výroba“), od které se odvíjí již konečná reálná budoucí výroba, sloupec ‚T‘. Pak je nutné zohlednit jak určující výrobu (sloupec ‚R‘), tak nezbytnou potřebu vyrobit takové množství, aby byl minimálně splněn sloupec ‚U‘, nutnost vyrobit.

Když *Evolver* rozhodne o výrobě položky (ano/ne=1/0, sloupec ‚R‘), pak se bude vyrábět (1 = ano). Jinak, když je nedostatek potřebného počtu výrobku i přes to, že je něco na skladě a například současně jistá část vyrobena, pak rovněž „ano“, bude se vyrábět.

Sloupec ‚V‘ představuje hodnotu, kterou je nutno ještě vyrobit, přesněji odečtení od výroby určené programem *Evolver* a množstvím na skladě.

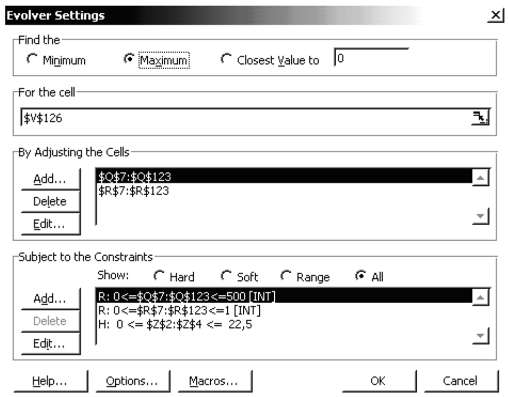
Když potřeba nepřevyšuje aktuální množství, pak není třeba vyrábět (nulový počet potřebných výrobků – 0), v opačném případě je nezbytné vyrobit minimálně právě onen nedostatek.

Protože zde jde o cyklické odkazy, vzájemně závislé podmínky, musíme v *Excelu* zvýšit počet iterací výpočtu na tři iterace v nabídce „Nástroje–Možnosti…“.

Sloupce ‚W‘, ‚X‘ a ‚Y‘ obsahují násobky párů počtu vyrobených kusů na jednom ze tří druhů lisů pod označením jejich počtu 15, 5 a 1. Počet minut vulkanizace u každé položky je uveden ve sloupci ‚C‘. Sloupec ‚AA‘ představuje množství na skladě na začátku po odečtení zakázek (sloupec ‚U‘) od výroby a skladových zásob. Jednoduchý odpočet samozřejmě nestačí, v případě nesplnění potřebné výroby by hodnota byla záporná, což je zjevný nesmysl. Problém vyřešíme snadnou podmínkou, když je rozdíl záporný, napiš nulu. V 2. až 4. řádku ve sloupci ‚U‘ a ‚V‘ je součet hodin vulkanizací všech výrobků, jinak řečeno, čas provozu jednotlivých lisů. Toto vše vykonáme pro celý měsíc září 2004. Soboty a neděle mají pouze dvanáctihodinové směny.

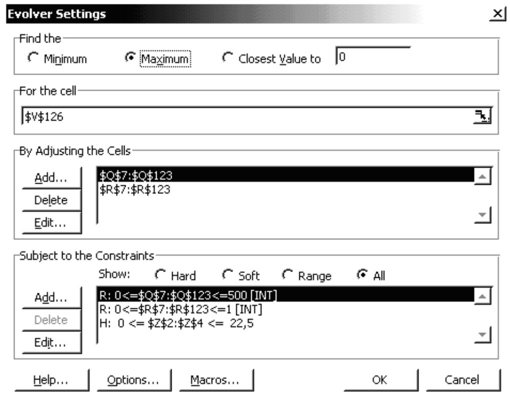
2.3. *Evolver*, software pro řešení složitých problémů genetickými algoritmy

Nyní přecházíme na konkrétní aplikaci genetických algoritmů, viz Obrázek 3. V kapitole, která zde v článku není obsažena, jsme vypočetli pravděpodobnou následující hodnotu u každého výrobku nejvhodnějším modelem a tuto hodnotu vložíme do posledního sloupce, od kterého odečteme veškerou uvažovanou výrobu. Výsledné hodnotě můžeme určit důležitost prostým váhovým koeficientem, stejně tak u každého výrobku v každém měsíci.



Obrázek 3. Základní zobrazení programu *Evolver*.

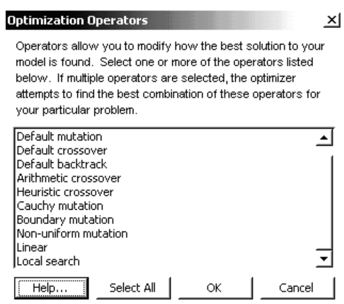
Do políčka „For the cell“ vložíme funkci, kterou chceme v průběhu optimalizace minimalizovat, maximalizovat nebo přiblížit konkrétní hodnotě, v našem případě se budeme snažit maximalizovat celkový výstup provozu, suma všech produktů. K tomu je ale nutné také určit buňky, které bude algoritmus upravovat pro dosažení nejlepší kombinace pro výstup a stanovit podmínky a omezení. Buňky pro úpravu se vkládají do políčka „by adjusting the cells“, viz Obrázek 3, podmínky a omezení se vkládají přes políčka „Subject to the Constraints“. Sloupec ‚R‘ (vyrobit tento výrobek – ANO/NE) bude obsahovat pouze binární hodnoty, tedy omezení na celá čísla v rozsahu $\langle 0; 1\rangle$ a u sloupce ‚Q‘ (velikost výroby), který označíme jako sloupec pro úpravu, nastavíme také celá čísla, ale s hranicí například $\langle 0; 500\rangle$ a metodu pro výpočet („Solving method“) v obou případech „Recipe“, viz Obrázek 4. Program nám sice nabízí i jiné metody, ale pro naše účely je vhodná pouze tato.



Obrázek 4. Nastavení parametrů nezávisle proměnné.

Mutační konstantu nastavíme na nízkou hodnotu 0,02, což je v praxi přibližně velmi častá hodnota, a práh křížení na hodnotu 0,9, která určuje pravděpodobnost, že se gen chromozomu nezmění, viz také Obrázek 4. Mutační konstanta může být zvolena i „auto“, tzn. automatické zvolení nejlepší hodnoty a při konvergenci poměrné zvyšování, my však, jelikož známe povahu problému, budeme konstantu během procesu řešení upravovat ručně.

Software není omezen jen na pouhé základní algoritmy pro proces selekce–křížení–mutace, poskytuje i modifikované algoritmy vhodnější pro specifické problémy, viz Obrázek 5.



Obrázek 5. Modifikace algoritmů.

Kromě základních („default“) algoritmů, je možné použít i následující postupy (algoritmy):

Arithmetic crossover: aritmetická kombinace dvou rodičů,

Heuristic crossover: využívá pozitivního trendu změny jedinců,

Cauchy mutation: malé změny při mutaci, občas větší,

Boundary mutation: pro rychlejší nalezení optima dává hraničním hodnotám vyšší váhy,

Non-uniform mutation: čím je vyšší počet pokusů, tím snižuje mutační konstantu pro konečné doladění,

Linear: vhodné pro lineární problémy,

Local search: algoritmus se zaměří na nejbližší minimum funkce a dále dolaďuje.

Pokud zvolíme všechny uvedené algoritmy, to program dovoluje, vybírá při procesu řešení nejlepší možnost. Pro náš případ, který opravdu není složitý, vybrat vše můžeme. Z počátku *Evolver* pravděpodobně nezvolí Cauchyho mutaci, ale pro doladění a předčasnou konvergenci je algoritmus vhodný.

Vyřešení problému na jeden den je ve velké většině vyřešen už po dvou minutách, a to i přesto, že byla zadána potřebná výroba využívající lisy z 90 % a současně nebyla algoritmu poskytnuta pomoc v podobě stanovení potřebné výroby již na počátku hledání řešení. To je velmi dobrý výsledek.

3. Problém obchodního cestujícího

Přestože název se může zdát zavádějící, algoritmus řešící tzv. problém obchodního cestujícího je možné užít i pro řešení našeho problému stanovení vhodného pořadí výrobků při penalizaci z přechodu mezi výrobky ve výrobě, kde například penalizace přechodu z výrobku A na výrobek B může být odlišná pro opačné pořadí (přechod z výrobku B na výrobek A). Koneckonců se toto specifikum může objevit i v realitě u problému obchodního cestujícího z důvodu jednosměrných ulic. Přestože zde jde pouze o neukončený okruh, nás tento detail neodradí od uvedeného názvu kapitoly. Vytvoření Hamiltonovy kružnice při pertubaci permutací je totiž ve vybraném algoritmu velmi jednoduché, spočívá pouze v kratičkém doplnění algoritmu o přechod mezi prvním a posledním prvkem, taková úprava by neměla trvat více než pár desítek sekund. Řízenou pertubaci permutací tedy použijeme při řešení problému.

Úkol a řešení

Úkolem bylo stanovit optimální pořadí vkládání receptur. Pro stanovení pořadí výroby je nejprve nutné získat data z provozu, která obsahují čas přechodu z receptury na recepturu u stroje zvaného TROESTER, viz Obrázek 6 nebo „Chalandra“, viz Obrázek 1.

zavádění rozměrů 1.4. - 4.4. 2002 na VL TROESTER 200/75

1.4.2002			
č.	změna rozměru	čas	doba (min)
1.	710 na 716	0:04	1
2.	716 na 723	1:34	2
3.	723 na 715	2:41	2
4.	715 na 712	3:18	2
5.	712 na 016	5:10	42
6.	016 na 299	6:32	21
7.	299 na 007	7:19	11
8.	007 na 047	8:10	11
9.	047 na 284	8:46	10
10.	284 na 046	9:30	20
11.	046 na 131	11:12	5
12.	131 na 335	11:22	0
13.	335 na 716	22:07	10
			10,5
2.4.2002			

Obrázek 6. Přechodové časy mezi recepturami.

Bohužel, podnik má k dispozici jen asi 50 přechodů, a jelikož v září 2004 bylo třeba vyrobit 117 produktů se 105 recepturami, není možné vytvořit kompletní řešení. Přesto jsme navrhli konkrétní řešení určení pořadí zpracování druhů surovin (receptur), protože tento problém je považován za velmi potřebný k optimalizaci. V Matlabu byl vytvořen algoritmus pro řešení problému obchodního cestujícího genetickými algoritmy, abychom potvrdili název kapitoly. Přesný postup řešení:

3.1. Stanovení konstant

Potřebné konstanty jsou: velikost populace, mutační konstanta a počet generací.

3.2. Generace jedinců

Za minimálně účinné množství chromozomů, jedinců, se považuje dvanáct (dvanáctičlenná populace), přesto pro jistotu zvolíme populaci o sto jedincích. Receptury byly označeny čísly od 1 do 105 a následně vytvořeno sto náhodných permutací 105 receptur.

3.3. Křížení

V našem specifickém problému se jedná o přenesení náhodné části jedince na druhého jedince. Protože v případě přenesení pouze jednoho genu se jedná v podstatě o krok 3.4., mutaci je dobré nastavit tak, aby náhodná část byla posloupnost více hodnot chromozomu. Křížení je velmi dobré pro přenášení nejlepších vlastností na jiné jedince.

3.4. Mutace

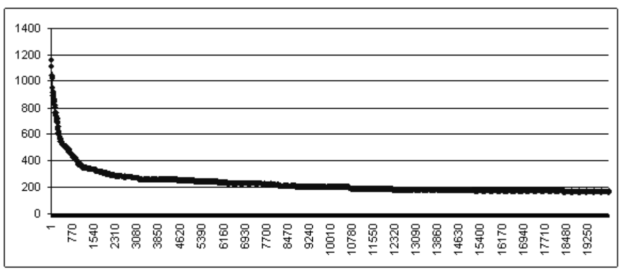
Literatura nabízí dva základní druhy mutace, a to „Exchange mutation“ a „Shift mutation“. Jak názvy napovídají, v prvním případě se jedná o výměnu libovolného počtu genů v chromozomu a v druhém případě jde o vyjmutí zvolených genů a umístění na libovolné pořadí v chromozomu. Zvolili jsme „Shift mutation“, z důvodu menšího „rozházení“ při novém přeskupení. Zatímco v posuvné mutaci při přesunutí jednoho genu dojde ke třem změnám, ve výměnné mutaci dojde ke čtyřem.

3.5. Výběr lepšího jedince

Výběr lepšího jedince je ve své podstatě obecně načtení celé funkce pro výpočet konečné hodnoty. Obvykle to bývá časově nejnáročnější operace, v našem případě to však není pravda.

3.6. Konečné řešení

Pro přijatelné řešení stačilo pouhých 10 000 generací, to představuje pět minut na středně výkonném počítači. My jsme zvolili 20 000 generací, výsledek, viz Obrázek 7.



Obrázek 7. Průběh nejlepších hodnot při řešení problému obchodního cestujícího.

Výsledná hodnota 167 minut ztrát při výměně 105 receptur, tzn. 104 změn, je s porovnáním s počáteční náhodnou permutací pořadí receptur (1162 minut) velmi dobrý výsledek. Pro bližší představu původní průměrný čas na přenastavení TROESTERU byl $1162/104 = 11{,}17$ minut. V nově uspořádaném pořadí jsme během desíti minut dospěli k potřebě pouhé 1,61 minuty na přenastavení stroje. Výsledek představuje $7\times$ lepší využití času, což při velmi častých změnách může v takové společnosti jako Mitas, a. s., představovat statisícové částky ročních úspor v nákladech na zaměstnance, protože mohou ušetřený čas upotřebit na jiné aktivity, které by vykonávali jiní pracovníci.

Volba mutační konstanty se ukázala jako ne příliš významná, jen je důležité, aby nenabyly příliš vysokých hodnot, nejlépe do 0,3. Nastavení mutační konstanty na hodnotu 50 % a výše není možné, to vychází zejména z povahy algoritmu, který problém řešit neumí. Z praktického hlediska nemá ani smysl algoritmus upravovat, protože výsledek po výměně více jak 50 % hodnot bude mít podobný užitek jako vytvoření nového, náhodně složeného jedince. Vyřešení problému by mělo značně malý mezní užitek. Výměna méně hodnot bude jistě efektivnější zejména v konečném doladění.

4. Závěr

Pro aplikaci genetických algoritmů bylo rozhodnuto z důvodu složitosti výrobního procesu přípravy motoplášťů, který dokáže optimalizovat snad jen řízená metoda pokus-omyl. Genetické algoritmy právě tuto metodu více než vylepšily a jsou snad jediným uvažovaným nástrojem pro řešení velmi složitých problémů s velkým množstvím proměnných.

Po vytvoření predikce výběrem nejvhodnější matematické či statistické metody a analýze výrobního procesu byl v *MS Excelu* zkonstruován model, který v potřebném směru odpovídal skutečné povaze výroby. Efektivnost výroby lze interpretovat jako množství vyrobených kusů s maximálním využití kapacit, proto zde byla jako účelová funkce vybrána rovnice, jejíž výsledek určoval minimálně požadované množství vyrobených motoplášťů. V článku popisujeme přesněji způsob řešení a využitý nástroj *Evolver* od společnosti *Palisade*.

Součástí úkolu byly samozřejmě mnohé dílčí podúkoly, mezi které patřil problém určení vhodného pořadí receptur při vkládání do výroby. Úkol byl rovněž řešen pomocí evolučního algoritmu, ale typizovaného pro „Problém obchodního cestujícího" v Matlabu.

Literatura

[1] Dostál, P. (2002) *Moderní metody ekonomických analýz: finanční kybernetika.* Zlín: Univerzita Tomáše Bati ve Zlíně, Fakulta managementu a ekonomiky. 110 s. ISBN 80-7138-075-8.

[2] Müller, T (2001). *Interaktivní tvorba rozvrhu.* Praha: Univerzita Karlova v Praze, Matematicko-fyzikální fakulta. 77 s. Bez ISBN.

[3] Zelinka, I (2002). *Umělá inteligence v problémech globální optimalizace.* Praha: BEN – technická literatura. 192 s. ISBN 80-7300-069-5.

INFORMACE O VZNIKU NÁRODNÍ SKUPINY ISCB V ČESKÉ REPUBLICE

Zdeněk Valenta

Adresa: EuroMISE Centrum, Ústav informatiky AV ČR, Pod Vodárenskou věží 2, 182 07 Praha

E-mail: `valenta@euromise.cz`

Po téměř rok trvající přípravné fázi byla dne 11. září 2006 rozhodnutím Ministerstva vnitra ČR s č.j. VS/1-1/65094/06-R v České republice zaregistrována *Mezinárodní společnost pro klinickou biostatistiku v České republice, o. s*, jako občanské sdružení (IČO 270 42 413). Společnost nabízí svým členům platformu vzájemné komunikace, sdílení poznatků a výměnu zkušeností na poli klinické biostatistiky, resp. v oblasti statistických aplikací v biomedicíně obecně.

Na mezinárodním poli je společnost integrální součástí *International Society for Clinical Biostatistics* (ISCB, http://www.iscb.info) jako její národní skupina. V současné době existují v rámci ISCB ještě tři další národní skupiny, a to v Polsku, Maďarsku a Rumunsku. Na úrovni mateřské ISCB byl vznik národní skupiny ISCB v České republice schválen hlasováním plnoprávných členů ISCB, jehož pozitivní výsledek byl vyhlášen dne 28. února 2006. S ohledem na stanovy ISCB byl ovšem vznik národní skupiny ISCB v České republice oficiálně potvrzen jejím současným prezidentem, prof. Johnem R. Whiteheadem z Univerzity v Readingu ve Velké Británii, až se šestiměsíčním zpožděním, tj. 28. srpna 2006, na výroční konferenci ISCB v Ženevě ve Švýcarsku.

Členové národní skupiny ISCB v České republice jsou plnoprávnými členy ISCB, jsou jim tudíž i poskytovány členské výhody na akcích ISCB. Přitom roční členský poplatek činí v současné době pouze 200 Kč (namísto 40 euro). K dnešnímu dni čítá společnost 34 plnoprávných členů.

Cíle činnosti národní skupiny mají vědecko-výzkumný a vzdělávací charakter. K hlavním plánovaným aktivitám patří kromě zprostředkování informací z mateřské ISCB především organizace odborných setkání, seminářů, krátkých kurzů pro biostatistiky, primárně zaměřených na aplikace statistických metod v klinické medicíně.

Na seminářích ISCB ČR jsme doposud uvítali Dr. Mansmanna a Dr. Komárka. Dr. Ulrich Mansmann, ředitel Ústavu zpracování medicínských informací, biometrie a epidemiologie Lékařské fakulty Ludwig-Maxmillian univerzity v Mnichově, Německo, přednesl přednášku s názvem *From technology* *driven research to hypothesis based questioning: microarrays in molecular medicine* s podtitulem *Complex Biomedical Data: Challenges to improve our statistical practice.* Na posledním semináři si členové společnosti vyslechli přednášku Dr. Arnošta Komárka z Katolické univerzity v Leuven, Belgie, s názvem *Semiparametrické regresní modely pro komplikovaná data o době do výskytu sledované události – MCMC pohled.* Členům společnosti jsou přes Internet zpřístupněny videozáznamy sdělení přednesených na seminářích.

ISCB pořádá každoročně výroční mezinárodní konferenci společnosti, která je prestižním podnikem v oboru klinické biostatistiky. V loňském roce se tato konference konala v Maďarském Szegedu, letos v Ženevě ve Švýcarsku, v příštím roce bude hostitelem Alexandroupolis na severovýchodě Řecka (29.7. – 2. 8. 2007) a v roce 2008 jím bude dánská Kodaň (termín předběžně 17. – 21. 8. 2008).

Národní skupina ISCB v České republice se uchází o možnost organizování 30. výroční konference ISCB v roce 2009 v České republice, kterou připravuje v součinnosti s konferenční agenturou *Guarant International, s.r.o.* Národní skupina již obdržela předběžný souhlas členů výkonného výboru ISCB s konáním této konference v Praze, a to v prostorách Vysoké školy ekonomické v Praze (Rajská budova). Předběžným termínem je konec srpna, resp. začátek září 2009.

Zájemci o členství v Mezinárodní společnosti pro klinickou biostatistiku v České republice, o. s., se mohou obrátit na kteréhokoliv ze zástupců společnosti, kterými jsou Zdeněk Valenta (předseda, valenta@euromise.cz), Marek Malý (zástupce předsedy, mmaly@szu.cz) a Věra Lánská (pokladník, vela@medicon.cz).

ZPRÁVA Z RADY VĚDECKÝCH SPOLEČNOSTÍ

Jan Picek

Adresa: Technická Univerzita v Liberci

E-mail: `jan.picek@vslib.cz`

Dne 29.listopadu jsem se zúčastnil plenární schůze Rady vědeckých společností ČR, což je zastřešující organizace vědeckých společností z přírodovědných, lékařských, společenskovědních a technických oborů. Mezi ně které patří i naše Česká statistická společnost. V současné době Rada koordinuje 72 společností. Tvoří ji zástupci jednotlivých společností, kteří volí devítičlenný výkonný výbor v čele s předsedou.

Jednání schůze pozdravil předseda akademie prof. Pačes, který zdůraznil význam vědeckých společností a apeloval na propagaci činnosti jednotlivých společností. Poté vystoupil předseda Rady prof. Hána a shrnul činnost za uplynulé funkční období. Objasňoval také problematiku získávání dotací na činnost a zdůrazňoval jak je důležitá popularizaci vědy a získávání nových mladých vědců.

Nejdůležitějším bodem schůze byly volby výboru. Na předsedu kandidoval opět Prof. MUDr. Ivo Hána, CSc. z České imunologické společnosti. Dalších 9 členů výboru tvoří zástupci tří sekcí (tři v každé): živá příroda, neživá příroda, společenské vědy. Česká statistická společnost je řazena do třetí sekce. Na tyto místa bylo navrženo celkem 12 kandidátů. V tajném hlasování byl velkou většinou přítomných delegátů (schůze se zúčastnilo 42 společností) opět zvolen předsedou prof. Hána a proti minulému období jen mírně obměněný devítičlenný výbor. Složení nového výboru by se mělo brzo objevit na `http://www.cas.cz/rvs/`, kde lze též nalézt další informace o Radě.

Na programu schůze byla též velmi zajímavá přednáška dr. Folty o historii vědeckých společností. V následné diskusi zástupci historických společností upozorňovali na důležitost schraňování materiálů významných vědců, důležitých hlášení apod. pro budoucí bádání. Vzniklo doporučení, aby si každá společnost založila archív. Archiv akademie nabízí odbornou pomoc.

LETNÍ ŠKOLA DATASTAT 06

Marie Budíková

Adresa: Masarykova Univerzita v Brně

E-mail: `budikova@math.muni.cz`

Po tříleté přestávce se letos opět konala letní škola DATASTAT, a to v termínu 4. – 7. 9. 2006. Již tradičně ji pořádala Katedra aplikované matematiky Přírodovědecké fakulty Masarykovy univerzity a Katedra aplikované matematiky a informatiky Ekonomicko-správní fakulty Masarykovy univerzity, k nimž se tentokrát přidalo nově zřízené Centrum Jaroslava Hájka pro aplikovanou a teoretickou statistiku. Organizační výbor ve složení prof. Horová, dr. Bauer, dr. Budíková a dr. Lepka vybral pro místo konání hotel Valáškův grunt v Kozově. Kozov je součástí obce Bouzov a nachází se v nejzápadnějším výběžku olomouckého okresu na hranici Čech a Moravy. Je to velmi klidné místo v lesnaté krajině v údolí říčky Třebůvky.

Vydatnou pomoc organizačnímu výboru s přípravou konferenčních materiálů poskytla paní Radka Paliánová, sekretářka katedry aplikované matematiky Přírodovědecké fakulty Masarykovy univerzity, za což jí patří dík a uznání, stejně jako Pavlu Budíkovi, studentu Fakulty informatiky Masarykovy univerzity, který vytvořil webovou stránku letní školy a průběžně se staral o její aktualizaci.

Na letní školu přijelo téměř šest desítek účastníků. Jednání probíhalo od pondělí do čtvrtka po pět půldnů přibližně po čtyřech hodinách. Příspěvky pokrývaly velmi široké spektrum problémů. Byly zde zastoupeny následující oblasti: lineární modely a jejich různé modifikace, vícerozměrné metody, evoluční algoritmy, stochastické programování, různé metody odhadu parametrů pravděpodobnostních rozložení, odhady ROC křivek a AUC, wavelety, vlastnosti diskrétních rozložení, konverze hlasu, optimalizace šablon, neuronové sítě, rozhodování za rizika, aplikace statistiky v medicíně, ekonomii, klimatologii, porovnání různých statistických programových systémů.

Pro doktorandy vytvořili organizátoři zvláštní sekci. Příspěvky v této sekci se zabývaly jádrovým vyhlazováním, zpracováním medicínských dat různými statistickými postupy, lineárními modely, neparametrickými testy, strojovým učením a ROC křivkami. Většina doktorandů si připravila postery a v průběhu pouhých pěti minut (překročení času nebylo tolerováno) se snažila seznámit posluchače s hlavními výsledky své práce. Jednalo se o velmi pěkně připravené prezentace, které byly odměňovány spontánním potleskem. Příznivé přijetí doktorandských příspěvků statistickou komunitou jistě povzbudí jejich autory k další vědecké a odborné činnosti.

Středeční odpoledne bylo věnováno velmi atraktivnímu výletu na hrad Bouzov. Na hrad bylo možno se vydat pěšky po turistických cestách. Kdo využil této možnosti, ocenil podrobný popis cesty, který pečlivě připravil dr. Bauer. Prohlídka nádherného hradu patřila k nejhezčím zážitkům z DATASTATu. Po návratu na všechny čekala vynikající večeře v podobě pečeného selete. Zábava pokračovala ještě dlouho po večeři a samozřejmě se netýkala jenom statistiky …

Ve čtvrtek před polednem DATASTAT skončil společným fotografováním. Řada účastníků využila dopravy autobusem, který zajistili organizátoři. Stejně jako v pondělí vedla trasa přes Olomouc, kde mnozí vystoupili, aby se mohli zúčastnit jednání navazující konference ODAM 06, kterou pořádala Katedra matematické analýzy a aplikací matematiky Přírodovědecká fakulty Univerzity Palackého v Olomouci.

Organizátoři předpokládají, že sborník příspěvků z letní školy DATASTAT vyjde v průběhu roku 2007 v řadě publikací Folia Přírodovědecké fakulty Masarykovy univerzity.

Vážené kolegyně, vážení kolegové,

Valná hromada společnosti a s ní spojená volba nového výboru se bude konat ve čtvrtek 8. února 2007 od 13.00 hodin v zasedací síni Českého statistického úřadu, Na Padesátém 81, 100 00 Praha 10, 100 metrů od konečné metra A Skalka. Odbornou přednášku přednese předseda ČSÚ, Ing. Jan Fischer, CSc.

Výbor společnosti Vás tímto žádá o návrhy kandidátů na funkci předsedy a do výboru společnosti a těší se na setkání na Valném shromáždění.

S přáním mnoha úspěchů, štěstí a zdraví v roce 2007

redakce a výbor společnosti.

Informační Bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. ISSN 1210 – 8022. <http://www.statspol.cz/>
Předseda společnosti: Prof. RNDr. Jaromír Antoch, CSc., KPMS MFF UK Praha, Sokolovská 83, 186 75 Praha 8, e-mail: jaromir.antoch@mff.cuni.cz
Redakce: Doc. RNDr. Gejza Dohnal, CSc. a Ing. Pavel Stříž, e-mail: gejza.dohnal@fs.cvut.cz a striz@fame.utb.cz