STOCHASTICKÝ PŘÍSTUP V MODELOVÁNÍ ŘEŠENÍ NĚKTERÝCH TERORISTICKÝCH HROZEB

Úvod

Jedním z největších problémů současného světa je nebezpečí terorismu. Toto nebezpečí stále vzrůstá a zdá se, že lidstvo nemá zbraň, která by toto hrozbu zastavila. Situací se zabývají nejenom ozbrojené a výzvědné složky států, ale v posledních letech také světová vědecká komunita. Snaha o nalezení exaktního řešení je zcela evidentní. Jednotlivé státy takový výzkum jednoznačně podporují. Bohužel zatím je tato snaha silně omezována obtížemi, které ji provázejí už od samotné formulace problému. Ucelený model, který by řešil terorismus jako celek, zatím nebyl podán.

Různí autoři používají velmi různé, metodologicky často velmi vzdálené metody a přístupy k řešení. Je snaha využít buď analogii s podobně formulovanými problémy z jiných oblastí lidské činnosti, např. z ekonomiky, kde hotové modely již dlouho existují, fungují a přináší dobré výsledky. Úspěšný postup tímto směrem obvykle zastaví nebo alespoň omezí zjištění, že do výsledků zasahují takové aspekty, jako jsou psychologie, sociologie, náboženství apod. s krajně obtížnou, ne-li úplně nemožnou kvantifikací. K takto dosaženým výsledkům je potom nutné přistupovat s velkou obezřetností a tím se dostáváme do problému značné neurčitosti. Proto se někteří snaží uplatnit principy známé z pravděpodobnosti a statistiky nebo i jiných oborů, které s neurčitostí pracují.

Z historického hlediska používané metody dělíme na metody kvalitativní a metody kvantitativní. Podle dosud řečeného by se na teroristické modely hodily spíše ty první, protože bychom se tak vyhnuli nutnosti kvantifikovat. Nicméně, takto dostáváme obvykle pouze globální pohled na věc, resp. obecné zásady, které nedávají návod co konkrétně dělat. Kvantitativní metody se hned na začátku dostávají do problému s tím, že vstupní data jsou získána „kvalifikovaným odhadem“. Od té chvíle však mohou přinést přesné výsledky za použití exaktních metod. Kvalifikovaný odhad se stává použitelným, když data získáváme rozborem velkého množství nepřímých dat. Potom může fungovat zákon velkých čísel a s velkým množstvím dat si počítače snadno poradí. Proto kvantitativní metody představují pokročilejší stav výzkumů.

Je rovněž účelné třídit metody z hlediska rozsáhlosti a podrobnosti řešení. Publikovaná řešení se zabývají malými, drobnými problémy, které mají za cíl hlavně prověřit, zda takový postup může v boji proti terorismu nějak pomoci až po vybudování rozsáhlých systémů, jejichž aplikaci nezvládne jediný člověk a musí být dovedené až k řešení aplikovanému na počítačových sítích.

V tomto článku je řešena drobná úloha o detektoru lži, založená na použití podmíněné pravděpodobnosti, resp. na Bayesově vzorci.

Uvedený princip byl publikován již dříve [1] pro konkrétní zadání, tj. konkrétní hodnoty vstupujících parametrů. Zde ukážeme zobecnění postupu řešení na libovolné vstupní hodnoty,

které patří do smysluplných definičních oborů, a budeme diskutovat vliv jednotlivých parametrů na výsledné řešení.

Popis modelu

- Předpoklad 1: víme, že obvyklý poměr teroristů ve skupině všech podezřelých bývá 100.T%.
- Spolehlivost detektoru udává výrobce a činí 100.s%.
- Počet náhodně vybraných podezřelých je n.
- Předpoklad 2: Ten, kdo opravdu teroristou je, to vždy popře.
- Předpoklad 3: Detektor označí $n.s$ z nich za lháře.
- Otázka: Jaká je pravděpodobnost, že z n náhodně vybraných podezřelých podrobených testu na detektoru lži, alespoň jeden skutečně teroristou je?

Použitá označení jevů:

E ..... jev, že aspoň jeden podezřelý je terorista,
$\overline{E}$ ..... jev, že žádný z podezřelých terorista není,
G ..... jev, že detektor lži potvrdí $n.s$ lhářů a $n.(1–s)$ pravdomluvných.

Potom pravděpodobnost, že jedním výběrem skutečně vybereme teroristu, je podle prvního předpokladu T. Pravděpodobnost, že při n pokusech žádného teroristu nevybereme, je potom

$$P(\overline{E}) = (1-T)^n \,. \tag{1}$$

Nyní zkoumejme pravděpodobnost, že při jediném výběru vybereme lháře. Takový jev je průnikem dvou menších jevů, a to samotný výběr jedince z n podezřelých a výsledek detektoru lži. Při výběru máme čtyři možnosti:

a) Vybereme teroristu a detektor jej označí za lháře …pravděpodobnost je $T.s$.
b) Vybereme teroristu a detektor jej označí za pravdomluvného … pravděpodobnost je $T.(1\text{-}s)$
c) Vybereme neteroristu a detektor jej označí za lháře …pravděpodobnost je $(1\text{-}T).(1\text{-}s)$.
d) Vybereme neteroristu a detektor jej označí za pravdomluvného … pravděpodobnost je $(1\text{-}T).s$.

Otázka byla na výběr lháře. To nastane, když náš výběr byl správný (vybrali jsme opravdu teroristu) a detektor to potvrdí nebo jsme se zmýlili (vybrali jsme neteroristu) ale zmýlil se i detektor, tedy vybraný neterorista je označen za lháře. To znamená, že naší úloze vyhovují případy a) a c) . Tedy, pravděpodobnost, že při výběru (jediném pokusu) vybereme lháře, je

$$p = T.\, s + (1–T)\,.(1–s)\,. \tag{2}$$

Tento pokus budeme nyní n-krát opakovat za stejných podmínek. Jde tedy o Bernoulliovskou posloupnost nezávislých pokusů. Náhodná veličina, která ji popisuje, má binomické rozdělení pravděpodobnosti.

$$P(X = k) = \binom{n}{k} . p^k . (1-p)^{n-k} \quad \text{, kde } k \text{ nabývá hodnot z intervalu } k \in \langle 0; n\rangle\,.$$

Tedy, definujme náhodnou veličinu X, popisující počet osob označených jako lhář, která má binomické rozdělení s parametry n , $k = n.s$, p ve tvaru

$$P(G) = \mathrm{P}(X = n.s) = \binom{n}{n.s} . p^{n.s} . (1-p)^{n-n.s} . \qquad (3)$$

Protože s je pravděpodobnost, hodnota $n.s$ zůstává v rozmezí $\langle 0; n \rangle$, a tedy je vždy $n.s \leq n$. Současně ale je potřeba upozornit, že hodnota $n.s$ v uvedeném binomickém číslu nemusí být celočíselná a má proto smysl prověřit věrohodnost takového výrazu. V dalším se ovšem zmíněné binomické číslo vykrátí ve zlomku. Diskuzi povedeme později.

Nyní vyčísleme podmíněnou pravděpodobnost, že detektor označí $n.s$ z n podezřelých za lháře, za podmínky, že tam žádní nejsou. Náhodná veličina Y, která to popisuje, má opět binomické rozdělení s parametry n , $k = n.s$ a $p = 1 - s$, a proto platí

$$P(G|\overline{E}) = \mathrm{P}(Y = n.s) = \binom{n}{n.s} . (1-s)^{n.s} . s^{n.(1-s)} . \qquad (4)$$

Otázka zformulovaná v úloze je potom zodpovězena výpočtem podle Bayesovy věty:

$$P(E|G) = 1 - P(\overline{E}|G) = 1 - \frac{P(G|\overline{E}).P(\overline{E})}{P(G)} . \qquad (5)$$

Např. máme-li $T = 0{,}03$, $s = 0{,}8$, $n = 10$ (volba použitá v citované literatuře), dostaneme

$$P(\overline{E}) = (1-T)^n = (1 - 0{,}03)^{10} = 0{,}737424127$$

$$p = T.s + (1-T).(1-s) = 0{,}03.0{,}8 + 0{,}97.0{,}2 = 0{,}218$$

$$P(G) = P(X = 10.0{,}8) = \binom{10}{8} . 0{,}218^8 . (1 - 0{,}218)^2 = 0{,}000140371$$

$$P(G|\overline{E}) = P(Y = 8) = \binom{10}{8} . 0{,}2^8 . 0{,}8^2 = 0{,}000073728$$

$$P(E|G) = 1 - \frac{P(G|\overline{E}).P(\overline{E})}{P(G)} = 1 - \frac{0{,}000073728.0{,}737424127}{0{,}000140371} = 0{,}612678302$$

a jsme ve shodě s [1] . Volba parametrů n a s je ovšem taková, že $n.s$ je přirozené číslo.

V případě, že vstupní parametry nejsou voleny tak, aby číslo $n.s$ bylo přirozené, např. je-li $T = 0{,}03$, $s = 0{,}8$, $n = 6$, dostaneme

$$P(\overline{E}) = (1-T)^n = (1 - 0{,}03)^6 = 0{,}737424127$$

$$p = T.s + (1-T).(1-s) = 0{,}03.0{,}8 + 0{,}97.0{,}2 = 0{,}218$$

$$P(G) = P(X = 6.0{,}8) = P(X = 4{,}8) = \binom{6}{4{,}8} . 0{,}218^{4{,}8} . (1 - 0{,}218)^{1{,}2} = 0{,}000497095 . \binom{6}{4{,}8}$$

$$P(G|\overline{E}) = P(Y = 4{,}8) = \binom{6}{4{,}8} . 0{,}2^{4{,}8} . 0{,}8^{1{,}2} = 0{,}000337794 . \binom{6}{4{,}8}$$

$$P(E|G) = 1 - \frac{P(G|\overline{E}).P(\overline{E})}{P(G)} = 1 - \frac{0{,}000337794 . \binom{6}{4{,}8} . 0{,}737424127}{0{,}000497095 . \binom{6}{4{,}8}} = 0{,}498893503$$

Nyní se zabývejme problémem necelého čísla $n.s$.
Vyjádřeme binomické číslo pomocí faktoriálů

$$\binom{n}{n.s} = \frac{n!}{(n.s)!.(n-n.s)!} \tag{6}$$

Výraz na pravé straně (6) existuje pouze pro $n.s$ přirozené. Můžeme jej ale rozšířit tak, že faktoriály nahradíme Gama funkcí. Tím se definiční obor výrazu na pravé straně (6) upraví na všechna komplexní s výjimkou celých záporných a nuly. Pro řešený problém postačí, že $n.s$ může být reálné kladné.

$$\binom{n}{n.s} = \frac{\Gamma(n+1)}{\Gamma(n.s+1).\Gamma(n-n.s+1)} \tag{7}$$

Funkce Gama je definována integrálem,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$$

který může být nepříjemné řešit. Víme, že násobným použitím známého vztahu

$$\Gamma(\mathrm{x}+1) = \mathrm{x}.\Gamma(\mathrm{x}) \tag{8}$$

potřebujeme znát pouze její hodnoty v intervalu $\langle 0; 1 \rangle$.

Tam ji můžeme nahradit polynomickou interpolací, např. touto kvadratickou z hodnot faktoriálu v bodech 0, 1, 2)

$$\Gamma(\mathrm{x}+1) \approx 1 - 0{,}5\mathrm{x} + 0{,}5\mathrm{x}^2 \tag{9}$$

Pro přesnější práci lze použít aproximace vyšších stupňů, např. tuto 5-tého stupně

$$\Gamma(x+1) \approx 1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 \tag{10}$$

kde

$$a_1 = -0{,}57486 \ ; \ a_2 = 0{,}9512363 \ ; \ a_3 = -0{,}6998588 \ ;$$

$$a_4 = 0{,}4245549 \ ; \ a_5 = -0{,}1010678 \ ;$$

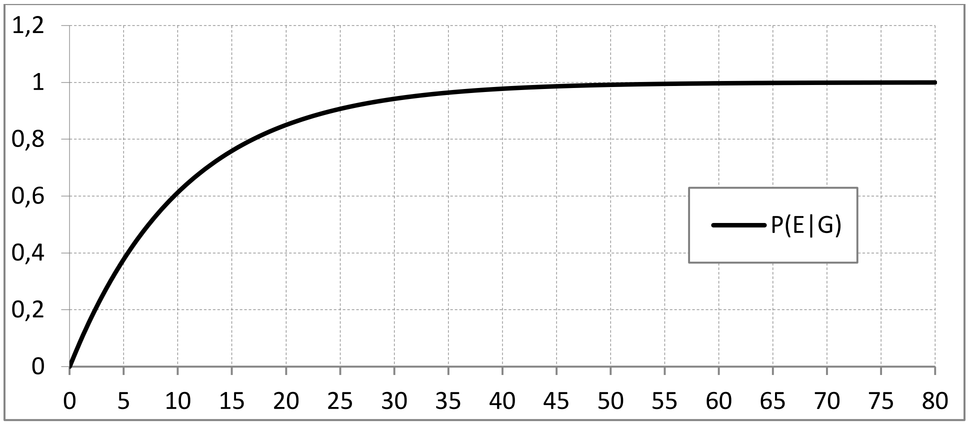
Průměrná odchylka obou zmíněných aproximací (9) a (10) od sebe navzájem je zhruba 0,005 , takže zpřesňující vliv aproximace 5-tého stupně není příliš významný.

Tímto jsme dokázali, že výraz $\binom{n}{n.s}$ pro přirozené *n* a *s* z intervalu $\langle 0; 1 \rangle$ vždy existuje. Zlomek, vyjadřující Bayesův vzorec, tedy má smysl.

Nyní zkoumejme vliv počtu podezřelých *n* na celkový výsledek.

Podržíme-li hodnoty ostatních parametrů konstantní, např. rovné těm, použitým v předchozích výpočtech, tedy *T* = 0,03 , *s* = 0,8 , dostaneme závislost podmíněné pravděpodobnosti $P(E|G)$ na počtu podezřelých *n* , což dokumentuje graf 1.

Původní otázka zněla: Jaká je pravděpodobnost, že z n náhodně vybraných podezřelých podrobených testu na detektoru lži, alespoň jeden skutečně teroristou je?

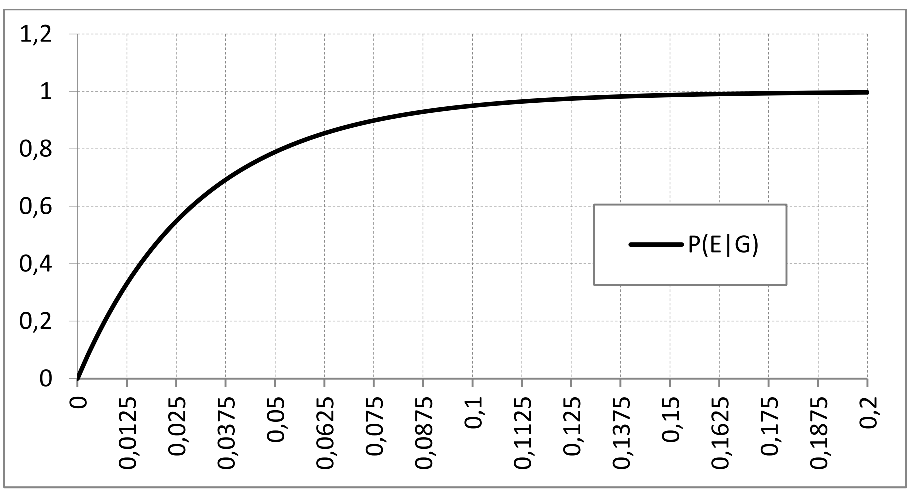


Graf 1. Závislost P(E|G) na počtu podezřelých

Z grafu je jasně patrné, že překročí-li počet podezřelých číslo asi 40, můžeme mezi nimi aspoň jednoho teroristu očekávat téměř s jistotou. Konkrétně pro n = 40 máme

$$P(E|G) = 1 - \frac{P(G|\overline{E}).P(\overline{E})}{P(G)} = 1 - \frac{0{,}2^{32}.0{,}8^{8}.0{,}97^{40}}{0{,}218^{32}.(1-0{,}218)^{8}} = 0{,}97749457$$

Podržíme-li $T = 0{,}03$, n = 10 konstantní, dostaneme závislost podmíněné pravděpodobnosti $P(E|G)$ na očekávaném poměru teroristů mezi podezřelými T , což dokumentuje graf 2.



Graf 2. Závislost P(E|G) na očekávaném poměru teroristů mezi podezřelými.

Literatura

[1] *Solution to problem 2, set 2, Math 3215*. School of Mathematics. Georgia Institute of Technology. [September 29, 2011]. Dostupné na internetu http://people.math.gatech.edu/

[2] KOVAŘÍK, P., SCHWARZ, R. *Matematický aparát řešení teroristických hrozeb*. In *Zborník príspevkov z V. medzinárodnej konferencie*, UMB Banská Bystrica, I. zväzok, s. 89-94. ISBN 978-80-557-0331-2.

STOCHASTIC APPROACH TO SOLUTIONS OF SOME TERRORIST THREATS MODELING

Trying to find the exact solution of such a problem is obvious. The quantitative methods are considered to be more useful than those qualitative ones. However, their weaknesses are imprecise input data obtained often only by a qualified estimate. The results thus obtained can then be approached with great caution and that brings us to the problem of considerable uncertainty.

In this article a minor problem of lie detector based on the use of conditional probabilities, respectively, on the Bayes formula is solved.

The principle of the solution has been published previously [1] for a specific input parameters. Here a generalization of the solution procedure for any input value with meaningful domains is shown. Later, the influence of various parameters on the resulting solution is discussed.

Key words: terrorist threats, conditional probability, Bayes formula, model parameters .

JEL Classification: C 02