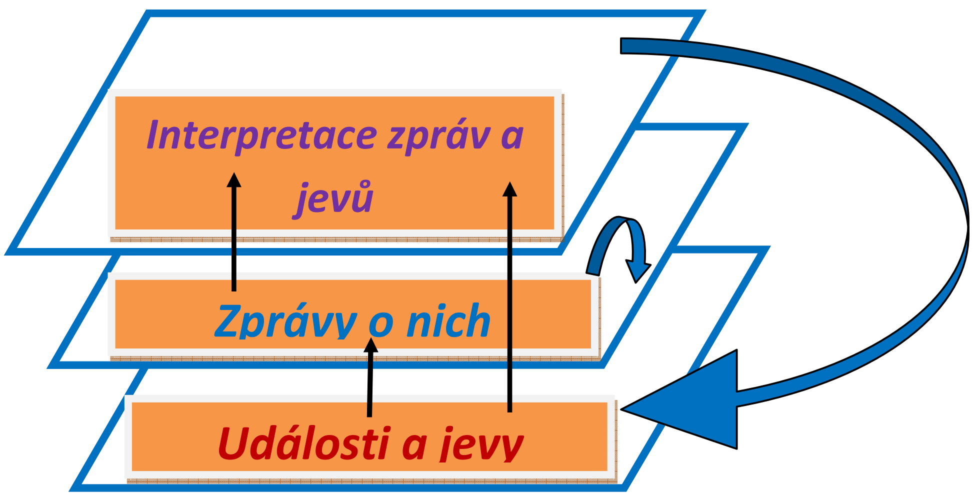
Teorie informace a aférologie, s příkladem pro případ ZČU

(některé poznámky k modelování hromadného sdělování)

František Vávra, květen 2011

Tři vrstvy obecného sdělování



Pět základních (lokálních) problémů současné doby

1. **Generování (pravdivých i jiných) zpráv nad prázdným jevem („reklama“, …).**
2. **Ze zpráv zjistit co se vlastně stalo, zprávy o zprávách.**
3. **Z interpretací zjistit na co vlastně reagují.**
4. **Interpretace generující nové události a jevy.**
5. **Rychlost (tzv. rychlé smyčky, „hazardy“), mnohost a dostupnost zpráv.**

Základní pojmy Shannonova teorie, diskrétní případ:

Konečné abecedy X, Y a nad nimi pravděpodobnosti elementárních jevů = prvků abecedy:

$$x \in X : p(x) \geq 0; \sum_{x \in X} p(x) = 1; [(x_1, x_2 \in X) \wedge (x_1 \neq x_2)] \Rightarrow p(x_1, x_2) = p(x_1 \wedge x_2) = 0$$

$$y \in Y : p(y) \geq 0; \sum_{y \in Y} p(y) = 1; [(y_1, y_2 \in Y) \wedge (y_1 \neq y_2)] \Rightarrow p(y_1, y_2) = p(y_1 \wedge y_2) = 0$$

Elementární informace:

1. $i(x) = i(p(x))$, **informační míra je funkcí pravděpodobnosti** (odvození z pravděpodobnosti obecně není nutné).

2. $x_1, x_2 \in X \wedge p(x_1) > p(x_2) \Rightarrow i(x_1) < i(x_2)$, ***koncepce překvapení***; pokud je přijata méně pravděpodobná zpráva, dostáváme více informace.

3. $x \in X; y \in X; x, y$ nezávislé $\Rightarrow i(x, y) = i(x) + i(y)$, **pozorování nezávislého jevu, přijetí nezávislé zprávy poskytuje úplně novou informaci.** Aditivita odvozená ze sčítání také není obecně nutná, může být pojata obecněji $x \in X; y \in X; x, y$ nezávislé $\Rightarrow i(x, y) = i(x) \oplus i(y)$, kde operace $\oplus$ je nějaká grupová (komutativní) operace, např. $i(x) \oplus i(y) = i(x) + i(y) - i(x)i(y)$.

4. **$i(p(x)) = i(p)$ je spojitou funkcí pravděpodobnosti.**

Z uvedených čtyř **základních** (před diskusí) požadavků lze odvodit, že $i(p(x)) = -\log_Z p(x); z > 1$. Důkaz [1], str. 46-49. Běžně je užíváno $i(p(x)) = -\log_2 p(x) = -lb[p(x)]$, často také $i(p(x)) = -\log_e p(x) = -\lg[p(x)]$.

Střední informace, míra informace zdroje – entropie:

$$H(X) = E\{i(x)\} = \sum_{x \in X} p(x) i(x) = \sum_{x \in X} p(x)(-\log p(x)) = E\{-\log p(x)\} = -\sum_{x \in X} p(x) \log p(x),$$

pokud je $p(x) = 0$, dodefinovává se $p(x) \log p(x) = 0$; vychází se z $\lim_{p \to 0_+} p \log p = 0$.

Platí: $0 \leq H(X) \leq \log|X|$. $H(X) = 0 \Leftrightarrow \exists x \in X; p(x) = 1$; $H(X) = \log|X| \Leftrightarrow \forall x \in X; p(x) = \frac{1}{|X|}$.

Odtud též název neurčitost informačního zdroje.

Koncepce přenosu, vazby nebo sdílení:

Dva, propojené informační zdroje s abecedami X, Y

$$H(X,Y) = E\{i(x,y)\} = -\sum_{x\in X, y\in Y} p(x,y)\log p(x,y)$$

$$H(X/Y) = E\{i(x/y)\} = -\sum_{x\in X, y\in Y} p(x,y)\log p(x/y)$$, **podmíněná entropie.**

$$H(Y/X) = E\{i(y/x)\} = -\sum_{x\in X, y\in Y} p(x,y)\log p(y/x)$$

Platí: $H(X) \geq H(X/Y)$. Důkaz: $H(X) - H(X/Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x) + \sum_{x\in X, y\in Y} p(x,y)\log p(x/y) =$

$$= \sum_{x\in X, y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = \sum_{x\in X, y\in Y} p(x,y)\left(-\log\frac{p(x)p(y)}{p(x,y)}\right) \geq$$

$$\geq -\log\left(\sum_{x\in X, y\in Y} p(x,y)\left(\frac{p(x)p(y)}{p(x,y)}\right)\right) = -\log(1) = 0$$

daná nerovnost je Jensenovou nerovností (viz: [3], str. 82-84).

$$I(X:Y) \overset{def}{=} H(X) - H(X/Y) = H(Y) - H(Y/X) = \sum_{x\in X, y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \geq 0$$, vazební, sdílená informace, informace.

Divergenční míry

Mějme abecedu elementárních zpráv (elementárních jevů) X a nad ní různá pravděpodobnostní rozdělení $P(x)$, $Q(x)$. Opět bude předpokládáno $\forall x \in X; P(x), Q(x) > 0$.

Divergence pravděpodobnostních modelů je zavedena klasicky $D(P\|Q) = \sum_{x\in X} P(x)\log\frac{P(x)}{Q(x)}$. Je to vlastně střední hodnota skóre testu jednoduché hypotézy (pozorování se řídí pravděpodobnostním rozdělením $P(x)$) proti jednoduché alternativě (pozorování se řídí pravděpodobnostním rozdělením $Q(x)$) za předpokladu, že platí hypotéza.

$D(P\|Q) \geq 0$. **Důkaz:** $D(P\|Q) = \sum_{x\in X} P(x)\log\frac{P(x)}{Q(x)} \geq \sum_{x\in X} P(x)\left(1 - \frac{Q(x)}{P(x)}\right) = 1 - 1 = 0$.

Jeffreysova J-divergence

Je zřejmé, že nemusí obecně platit $D(P\|Q) = D(Q\|P)$. Tedy divergence není symetrická. Proto se zavádí symetrická divergence $J(P,Q) = D(P\|Q) + D(Q\|P)$ (J-divergence, Jeffreysova divergence).

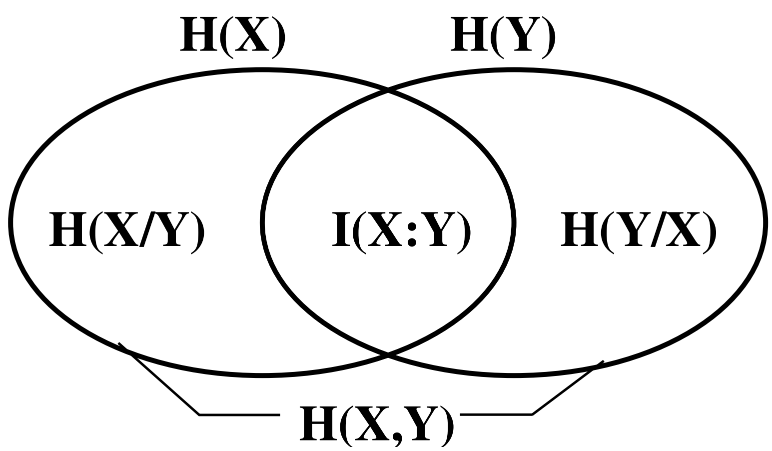
Platí:

H(X,Y) = H(Y,X), H(X,Y) = H(X/Y) + H(Y), H(X,Y) = H(Y/X) + H(X)

I(X:Y) = I(Y:X) , I(X:Y) = H(Y) - H(Y/X), I(X:Y) = H(X) - H(X/Y)

H(X,Y)=H(X)+H(Y)- I(X:Y)

Odtud interpretace pomocí Vénova diagramu:



Platí: X, Y nezávislé $\Leftrightarrow I(X:Y)=0$.

Popis sdělování (prezentace)

Mějme konkrétní zdroj informace s abecedou zpráv: $X=\{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ takové zprávy jsou věrně prezentovány příjemci s jeho abecedou $X=\{y_1, y_2, \ldots, y_i, \ldots, y_n\}$ s maticí přenosu $P(y_i / x_j) = \begin{matrix} 1 \Leftrightarrow i=j \\ 0 \Leftrightarrow i \neq j \end{matrix}; i, j = 1, \ldots, n$ a s pravděpodobnostmi výskytu jednotlivých zpráv $P(x_i) = p_i; i=1, \ldots, n$. Potom: $H(X) = H(Y) = -\sum_{i=1}^{n} p_i \lg p_i$, $I(X:Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} p(i,j) \lg \frac{p(i,j)}{p_i p_j} =$

$$= \sum_{i=1}^{n} p_i \lg \frac{p_i}{p_i p_i} = -\sum_{i=1}^{n} p_i \lg p_i = H(X).$$

„Štěpení“ zprávy při prezentaci

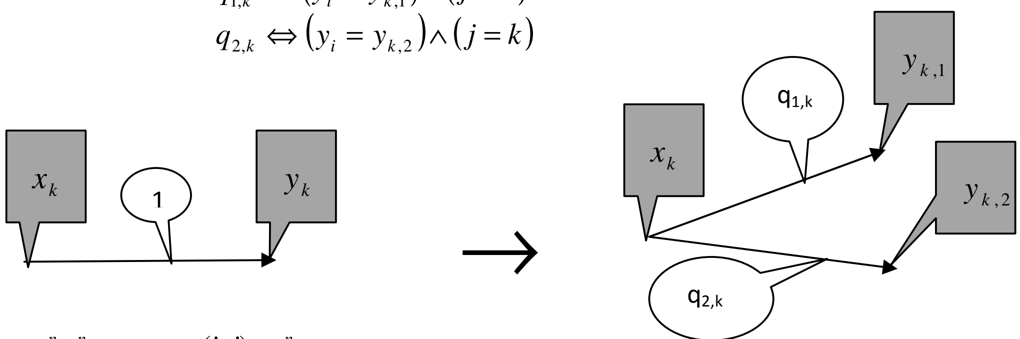
Nechť je nyní prezentační abeceda obohacena „rozštěpením“ jedné k-té zprávy $y_k \rightarrow \begin{matrix} y_{k,1} \\ y_{k,2} \end{matrix}$, potom:

$p_k = p(y_k) = p(x_k) = p(y_{k,1}) + p(y_{k,1})$.

Takové „rozštěpení“ má reálný podklad, původní zpráva x_k ≡“***stala se nehoda***“ je prezentována jako $y_{k,1}$ ≡“stala se ***tragická*** nehoda“. Taková prezentace samozřejmě vyvolá dojem (a i implicitní existenci) toho, že existuje jiná doplňková zpráva $y_{k,2}$ ≡“stala se ***nikoliv tragická*** nehoda“.

Matice přenosu má po takové „úpravě“ tvar:

$$P(y_i / x_j) = \begin{cases} 1 \Leftrightarrow i = j \neq k \\ 0 \Leftrightarrow (i \notin \{(k,1),(k,2)\}) \wedge (j \neq k) \\ q_{1,k} \Leftrightarrow (y_i = y_{k,1}) \wedge (j = k) \\ q_{2,k} \Leftrightarrow (y_i = y_{k,2}) \wedge (j = k) \end{cases} ;\ i, j = 1,\ldots,n;\ q_{1,k}, q_{2,k} > 0;\ q_{1,k} + q_{2,k} = 1$$



$$I(X:Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} p(i,j) \lg \frac{p(i,j)}{p_i p_j} = \sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg \frac{p_i}{p_i p_i} + q_{1,k} p_k \lg \frac{q_{1,k} p_k}{p_k (q_{1,k} p_k)} + q_{2,k} p_k \lg \frac{q_{2,k} p_k}{p_k (q_{2,k} p_k)} =$$

$$= \sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg \frac{p_i}{p_i p_i} + q_{1,k} p_k \lg \frac{1}{p_k} + q_{2,k} p_k \lg \frac{1}{p_k} = \sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg \frac{p_i}{p_i p_i} + (q_{1,k} + q_{2,k}) p_k \lg \frac{1}{p_k} = \sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg \frac{p_i}{p_i p_i} + p_k \lg \frac{1}{p_k} =$$

$$= -\sum_{i=1}^{n} p_i \lg p_i$$

Tedy: přenesen **je shodný informační obsah** zdroje jako v případě věrného přenosu. Na vstupní = zdrojové straně zůstává také původní stav $-\sum_{i=1}^{n} p_i \lg p_i = H(X)$ ale na výstupní straně

bude : $$H_{k,q_{1,k},q_{2,k}}(Y) = -\sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg p_i - q_{1,k} p_k \lg(q_{1,k} p_k) - q_{2,k} p_k \lg(q_{2,k} p_k) =$$

$$= -\sum_{\substack{i=1 \\ i \neq k}}^{n} p_i \lg p_i - q_{1,k} p_k \lg(q_{1,k}) - q_{1,k} p_k \lg(p_k) - q_{2,k} p_k \lg(q_{2,k}) - q_{2,k} p_k \lg(p_k) =$$

$$= -\sum_{i=1}^{n} p_i \lg p_i - q_{1,k} p_k \lg(q_{1,k}) - q_{2,k} p_k \lg(q_{2,k}) = -\sum_{i=1}^{n} p_i \lg p_i - p_k (q_{1,k} \lg(q_{1,k}) + q_{2,k} \lg(q_{2,k}))$$

Shrnuto: $H(Y) = -\sum_{i=1}^{n} p_i \lg p_i - p_k (q_{1,k} \lg(q_{1,k}) + q_{2,k} \lg(q_{2,k}))$ ≡ **střední poskytovaná informace (entropie) prezentační úrovně se zvětší o hodnotu** $-p_k (q_{1,k} \lg(q_{1,k}) + q_{2,k} \lg(q_{2,k}))$ **a ta nabývá svého maxima pro** $q_{1,k} = q_{2,k} = 1/2$ **s hodnotou** $p_k \lg(2)$ **.**

Ještě jednou shrnuto: $H_{k,q_{1,k},q_{2,k}}(Y) = H(Y) - p_k\left(q_{1,k}\lg\left(q_{1,k}\right) + q_{2,k}\lg\left(q_{2,k}\right)\right) \le H(Y) + p_k\lg(2)$

Celý postup měření je založen na následující „triviální“ nerovnosti. Nechť je zpráva x s pravděpodobností výskytu p „rozštěpena“ do dvou zpráv x_1, x_2 s pravděpodobnostmi výskytu $p_1, p_2;\ p_1 + p_2 = p;\ 0 < p_1, p_2, p < 1$. Pak platí:

$$-\left(p_1 + p_2\right)\log\left(p_1 + p_2\right) < -p_1\log\left(p_1\right) - p_2\log\left(p_2\right)$$

Odvození:

1. $-\left(p_1 + p_2\right)\log\left(p_1 + p_2\right) \quad \mapsto \quad -p_1\log\left(p_1\right) - p_2\log\left(p_2\right)$
2. $\log\left(p_1 + p_2\right)^{-\left(p_1 + p_2\right)} \quad \mapsto \quad \log\left(p_1^{\ -p_1}\right) + \log\left(p_2^{\ -p_2}\right)$
3. $\left(p_1 + p_2\right)^{-\left(p_1 + p_2\right)} \quad \mapsto \quad \left(p_1^{\ -p_1}\right)\left(p_2^{\ -p_2}\right)$
4. $\left(p_1 + p_2\right)^{-p_1}\left(p_1 + p_2\right)^{-p_2} \quad \mapsto \quad \left(p_1^{\ -p_1}\right)\left(p_2^{\ -p_2}\right)$
5. $\left(\dfrac{p_1 + p_2}{p_1}\right)^{-p_1}\left(\dfrac{p_1 + p_2}{p_2}\right)^{-p_2} \quad \mapsto \quad 1$
6. $\left(\dfrac{p_1}{p_1 + p_2}\right)^{p_1}\left(\dfrac{p_2}{p_1 + p_2}\right)^{p_2} \quad \mapsto \quad 1$ a zde je evidentní, že:
7. $\left(\dfrac{p_1}{p_1 + p_2}\right)^{p_1}\left(\dfrac{p_2}{p_1 + p_2}\right)^{p_2} \quad < \quad 1$

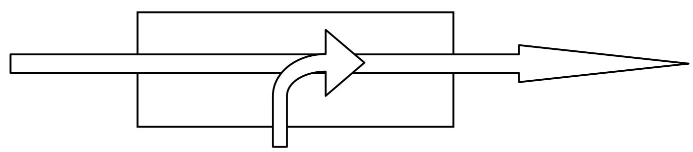
Důkaz: Opačným postupem od bodu 7. k bodu 1.

Uvedená nerovnost je specielním případem obecnější (Log-Sum inequality, [C])

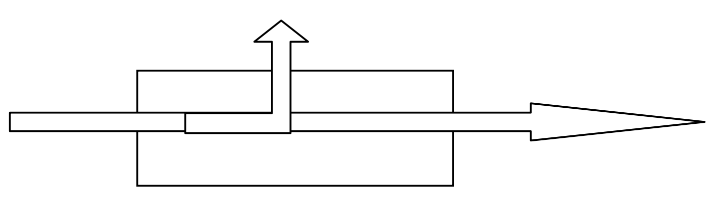
$$\sum_{i=1}^{n} a_i \log\frac{a_i}{b_i} \ge \left(\sum_{i=1}^{n} a_i\right)\log\frac{\left(\sum_{i=1}^{n} a_i\right)}{\left(\sum_{i=1}^{n} b_i\right)}; a_i, b_i \ge 0 \forall i = 1,\dots,n$$

Shrnuto:

„Štěpení zprávy“

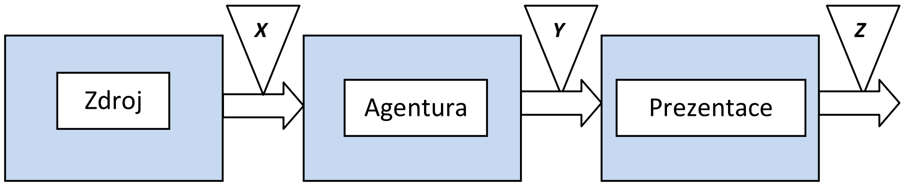


„Sloučení zpráv“



Zprávy z druhé ruky

Model - Markovská soustava.



Obecně platí: $P(x,y,z)=P(z/y,x)P(y,x)=P(z/y,x)P(y/x)P(x)$. Takovou přenosovou soustavu budeme nazývat markovskou pokud platí $P(z/y,x)=P(z/y)$ a bude to značeno $X \rightarrow Y \rightarrow Z$. Pro markovskou soustavu proto bude: $P(x,y,z)=P(z/y,x)P(y,x)=P(z/y,x)P(y/x)P(x)=P(z/y)P(y/x)P(x)$. Dále $P(x,z/y)=\dfrac{P(x,z,y)}{P(y)}=$

$$=\frac{P(z/y)P(y/x)P(x)}{P(y)}=\frac{P(z/y)P(x,y)}{P(y)}=P(z/y)P(x/y).$$

Shrnuto $P(x,z/y)=P(z/y)P(x/y)$. Odtud:

Definice: Abecedy X a Z budeme nazývat podmíněně nezávislé při daném Y, pokud platí: $P(x,z/y)=P(z/y)P(x/y) \quad \forall x \in X\ \forall y \in Y\ \forall z \in Z$. Daný pojem je vázán na pravděpodobnost nad danými abecedami. [C] str. 21-23.

Proti tomu je klasická nezávislost:

Definice: Abecedy X a Z budeme nazývat nezávislé, pokud platí: $P(x,z)=P(x)P(z) \quad \forall x \in X\ \forall z \in Z$. Opět je daný pojem vázán na pravděpodobnost nad danými abecedami.

Tvrzení: Podmíněná nezávislost neimplikuje nepodmíněnou nezávislost.

Tvrzení: Nepodmíněná nezávislost neimplikuje podmíněnou nezávislost.

Důkaz: je v příloze = příklady kdy hypoteticky předpokládané implikace neplatí.

Tvrzení: Je-li $X \rightarrow Y \rightarrow Z$, pak je i $Z \rightarrow Y \rightarrow X$

Důkaz: K tomu stačí dokázat $P(x/y,z)=P(x/y)$. Ale $P(x/y,z)=\dfrac{P(x,y,z)}{P(y,z)}=\dfrac{P(z/y)P(y/x)P(x)}{P(y,z)}=$

$$=\frac{P(z/y)P(y)P(y/x)P(x)}{P(y)P(y,z)}=\frac{P(y/x)P(x)}{P(y)}=\frac{P(x,y)}{P(y)}=P(x/y).$$

Definice: Podmíněná informace. $I(X:Y/Z)=\sum_{x,y,z \in XxYxZ} P(x,y,z)\log\dfrac{P(x,y/z)}{P(x/z)P(y/z)}$

Tvrzení: $I(X:Y/Z)\geq 0$

Důkaz: $\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y/z)}{P(x/z)P(y/z)}=\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\left[-\log\frac{P(x/z)P(y/z)}{P(x,y/z)}\right]\geq$

$$\geq -\log\sum_{x,y,z\in XxYxZ} P(x,y,z)\left[\frac{P(x/z)P(y/z)}{P(x,y/z)}\right]=-\log\sum_{x,y,z\in XxYxZ} P(x/z)P(y/z)P(z)=-\log\sum_{x,z\in XxYxZ} P(x/z)P(z)=-\log(1)=0.$$

Tvrzení: $I(X:Y/Z)=I((X,Z):(Y,Z))-H(Z)$, kde $I((X,Z):(Y,Z))=\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y,z)}{P(x,z)P(y,z)}$.

Důkaz: Prostým výpočtem.

Tvrzení: $X\to Y\to Z\Rightarrow I(X:Y/Z)=H(X/Z)-H(Y/Z)$

Důkaz: $X\to Y\to Z\Rightarrow P(z/y,x)=P(z/y)$ a pak $\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y/z)}{P(x/z)P(y/z)}=$

$$=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y,z)}{P(x/z)P(y/z)P(z)}=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(z/x,y)P(x,y)}{P(x/z)P(y/z)P(z)}=$$

$$=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(z/y)P(x,y)}{P(x/z)P(y/z)P(z)}=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y)}{P(x/z)P(y)}=$$

$$=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x/y)}{P(x/z)}=H(X/Z)-H(X/Y)$$

Důsledek: $X\to Y\to Z\Rightarrow H(X/Z)\geq H(Y/Z)$

Tvrzení: $X\to Y\to Z\Rightarrow I(X:Z/Y)=0$

Důkaz:

$$I(X:Z/Y)=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,z/y)}{P(x/y)P(z/y)}=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y,z)}{P(x/y)P(z/y)P(y)}=$$

$$=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x/y,z)P(y,z)}{P(x/y)P(y,z)}=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x/y)}{P(x/y)}=0$$

Tvrzení: $X\to Y\to Z\Rightarrow I(X:(Y,Z))=I(X:Y)$

Důkaz: $I(X:(Y,Z))=\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y,z)}{P(x)P(y,z)}=\sum\limits_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x/y,z)}{P(x)}=$

$$=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x/y)}{P(x)}=\sum_{x,y,z\in XxYxZ} P(x,y,z)\log\frac{P(x,y)}{P(x)P(y)}=I(X:Y)$$

Tvrzení: $X \to Y \to Z \Rightarrow I(X:(Y,Z)) = I(X:Y/Z) + I(X:Z)$

Důkaz:

$$I(X:(Y,Z)) = \sum_{x,y,z \in XxYxZ} P(x,y,z)\log\frac{P(x,y,z)}{P(x)P(y,z)} = \sum_{x,y,z \in XxYxZ} P(x,y,z)\log\frac{P(x,y/z)P(z)}{P(x)P(y/z)P(z)} =$$

$$= \sum_{x,y,z \in XxYxZ} P(x,y,z)\log\frac{P(x,y/z)}{P(x/z)P(y/z)}\frac{P(x/z)}{P(x)} = I(X:Y/Z) + \sum_{x,y,z \in XxYxZ} P(x,y,z)\log\frac{P(x,z)}{P(x)P(z)} =$$

$$= I(X:Y/Z) + I(X:Z)$$

Srovnáním obou předchozích tvrzení dostaneme: $X \to Y \to Z \Rightarrow I(X:Y) = I(X:Y/Z) + I(X:Z)$ a odtud: $X \to Y \to Z \Rightarrow I(X:Y) \geq I(X:Z)$ s rovností při podmíněné nezávislosti X a Y při daném Z, tj. $I(X:Y/Z) = 0 \Leftrightarrow \forall x,y,z \in XxYxZ;\ P(x,y/z) = P(x/z)P(y/z)$.

A obdobně $X \to Y \to Z \Rightarrow I(X:Y) \geq I(X:Y/Z)$

Námět: Dokažte, že předpoklad $X \to Y \to Z$ pro platnost nerovnosti $I(X:Y) \geq I(X:Y/Z)$ je podstatný.

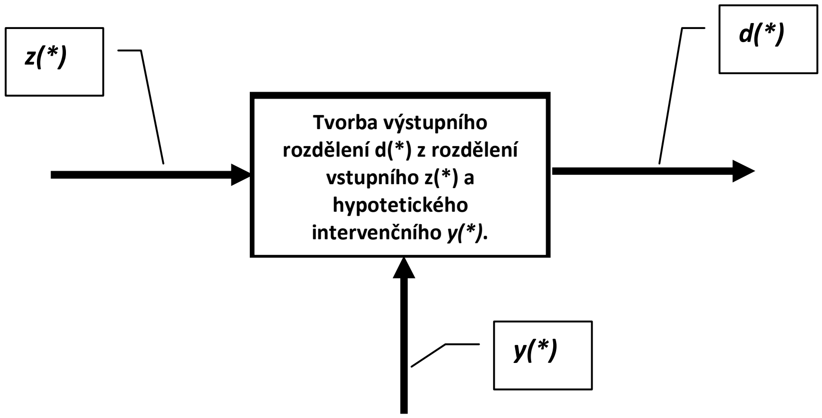
Tj. $\exists\ P(*,*,*)$ na $XxYxZ; I(X:Y) > I(X:Y/Z)$ a i $\exists\ P(*,*,*)$ na $XxYxZ; I(X:Y) < I(X:Y/Z)$.

První část je triviální, druhá nalezením příkladu.

Námět: Za jakých podmínek na $P(*,*,*)$ na $XxYxZ$ bude platit $I(X:Y) = I(X:Y/Z)$

Shrnuto: **Vložením prostředníka [umělý propojovací jev(y) nebo zpráva(y)] při sdělování, lze vyvolat „dojem“ souvislosti jevu a zprávy, které ve skutečnosti nemají nic společného. Nebo i naopak. Tento „dojem“ lze modelovat a lze ho i měřit.**

Model intervence



Předpokládáme, že daný model pracuje s konečnou množinou stavů, tu budeme nahrazovat čísly $\{1,2,...,n\}$ pro všechny vstupy a výstup, shodnou.

Pro analýzu budeme předpokládat model směsi rozdělení:

$$d(i) = (1-\alpha)z(i) + \alpha y(i); 0 < \alpha \leq 1$$

Odtud pro intervenční rozdělení $y(i)$ platí: $y(i) = \dfrac{d(i)-(1-\alpha)z(i)}{\alpha} = \dfrac{d(i)-z(i)}{\alpha} + z(i)$. Z toho, že $y(i); i = 1,...,n$ je pravděpodobnostní rozdělení pak pro koeficient směsi α (= síla intervence do změny) musí platit:
$0 \leq \dfrac{d(i)-(1-\alpha)z(i)}{\alpha} \leq 1$ a $\sum_{i=1}^{n} \dfrac{d(i)-(1-\alpha)z(i)}{\alpha} = 1$. Ke splnění poslední rovnosti je nutné a stačí $\alpha \neq 0$. Nadále bude zkoumána první dvojice nerovností: $0 \leq \dfrac{d(i)-(1-\alpha)z(i)}{\alpha}$ a $\dfrac{d(i)-(1-\alpha)z(i)}{\alpha} \leq 1$. Z první nerovnosti plyne: $\alpha \geq \dfrac{z(i)-d(i)}{z(i)}; i = 1,...,n$ a z druhé lze odvodit: $\alpha \geq \dfrac{d(i)-z(i)}{1-z(i)}; i = 1,...,n$, odtud plyne potřeba $0 < z(i) < 1$[1].

Odtud pro volbu síly intervence $\alpha \geq \max\left\{\dfrac{z(i)-d(i)}{z(i)}, \dfrac{d(i)-z(i)}{1-z(i)}; i = 1,...,n\right\}$

Je evidentní, že se jeden člen z dvojice $\dfrac{z(i)-d(i)}{z(i)}, \dfrac{d(i)-z(i)}{1-z(i)}$ pro každé *i* neuplatní (bude alespoň pro jedno *i* záporný a obecně nekladný). Pokud tomu tak nebude, jsou vstupní a výstupní rozdělení shodná a to vede na triviální řešení α=0 (tento případ bude nadále vynecháván). Aby existovalo „použitelné netriviální řešení pro α“ musí platit: $0 < \max\left\{\dfrac{z(i)-d(i)}{z(i)}, \dfrac{d(i)-z(i)}{1-z(i)}; i = 1,...,n\right\} < 1$.

[1] To může být např. zabezpečeno volbou metody odhadu vstupních pravděpodobností.

První nerovnost (kladnost) je splněna triviálně (za výše uvedeného předpokladu). Pro druhou budeme analyzovat dva možné případy:

1. $\max\left\{\frac{z(i)-d(i)}{z(i)},\frac{d(i)-z(i)}{1-z(i)};i=1,\ldots,n\right\}=\frac{z(k)-d(k)}{z(k)}$

2. $\max\left\{\frac{z(i)-d(i)}{z(i)},\frac{d(i)-z(i)}{1-z(i)};i=1,\ldots,n\right\}=\frac{d(l)-z(l)}{1-z(l)}$

První případ vede na podmínku $\frac{z(k)-d(k)}{z(k)}<1 \Leftrightarrow z(k)-d(k)<z(k) \Leftrightarrow -d(k)<0$. K jejímu splnění je nutné a stačí aby $d(k)>0$.

Druhý případ vede na podmínku $\frac{d(l)-z(l)}{1-z(l)}<1 \Leftrightarrow d(l)<1$. I tuto podmínku lze přijmout pro běžné situace.

Pokud by došlo k volbě: $\alpha=\max\left\{\frac{z(i)-d(i)}{z(i)},\frac{d(i)-z(i)}{1-z(i)};i=1,\ldots,n\right\}$, pak by ve výše uvedeném ***prvním případě*** platilo: $y(k)=\frac{d(k)-z(k)}{\alpha}+z(k)=\frac{d(k)-z(k)}{\frac{z(k)-d(k)}{z(k)}}+z(k)=0$. Při téže volbě by v ***druhém případě*** platilo: $y(l)=\frac{d(l)-z(l)}{\alpha}+z(l)=\frac{d(l)-z(l)}{\frac{d(l)-z(l)}{1-z(l)}}+z(l)=1$ a v tom případě pro $i\neq l \Rightarrow y(i)=\frac{d(i)-z(i)}{\frac{d(l)-z(l)}{1-z(l)}}+z(i)=0 \Leftrightarrow \frac{d(i)-z(i)}{d(l)-z(l)}\left(1-z(l)\right)+z(i)=0 \Leftrightarrow$

$$\Leftrightarrow \left(d(i)-z(i)\right)\left(1-z(l)\right)+z(i)\left(d(l)-z(l)\right)=0 \Leftrightarrow \frac{d(i)}{z(i)}=\frac{1-d(l)}{1-z(l)} \quad \Rightarrow i\neq l \Rightarrow d(i)=kz(i) \;\Rightarrow$$

$\Rightarrow \forall i\neq l; \frac{z(i)-d(i)}{z(i)}=\frac{z(i)-kz(i)}{z(i)}=1-k=1-\frac{1-d(l)}{1-z(l)}=\frac{d(l)-z(l)}{1-z(l)}$. Toto je důležitý případ, který říká: **pokud nastane druhý případ, nastává i první**. Proto $\alpha\geq\max\left\{\frac{z(i)-d(i)}{z(i)};i=1,\ldots,n\right\}$

A odtud, shrnuto: **Aby existovala netriviální intervence** $(\alpha\neq 0)\wedge(\alpha\neq 1)$**, musí být splněno** $\exists i=1,\ldots,n; d(i)\neq z(i)$ **a zároveň** $\forall=1,\ldots,n; d(i)>0$**.**

Tedy: Prakticky u všech dvojic zdrojů nad stejnou abecedou lze nalézt hypotetický, alespoň aproximativní, intervenční zdroj, který intervencí z jednoho rozdělení stavů (zpráv, událostí, ...) vytváří rozdělení druhé.

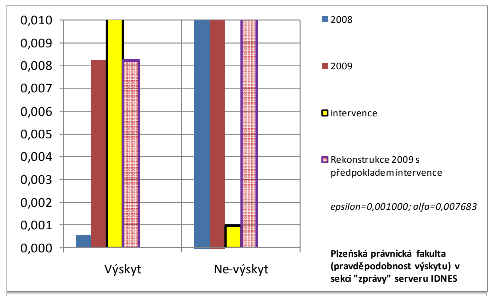
Hypotetická „spiklenecká intervence“

$$\mathbf{y(i)} = \boldsymbol{\varepsilon}; i = 1,\ldots,k-1, k+1,\ldots,n; \mathbf{y(k)} = \mathbf{1} - (\mathbf{n}-\mathbf{1})\boldsymbol{\varepsilon}; 0 < \varepsilon \leq 1/_{n-1}$$

Příklad takové rekonstrukce:

Plzeňská právnická fakulta (pravděpodobnost výskytu) v sekci "zprávy" serveru IDNES

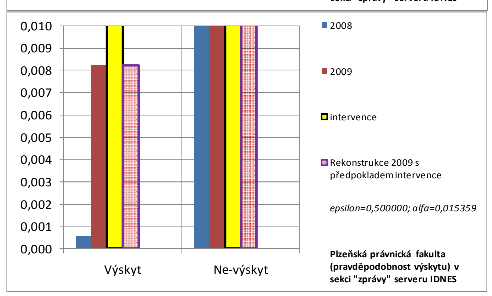
	2008	*2009*	*intervence*	Rekonstrukce 2009 s předpokladem intervence
Výskyt	0,00057	0,00824	0,99900	0,00824
Ne-výskyt	0,99943	0,99176	0,00100	0,99176



epsilon=	0,00100
alfa=	0,00768

Plzeňská právnická fakulta (pravděpodobnost výskytu) v sekci "zprávy" serveru IDNES

	2008	*2009*	*intervence*	Rekonstrukce 2009 s předpokladem intervence
Výskyt	0,00057	0,00824	0,50000	0,00824
Ne-výskyt	0,99943	0,99176	0,50000	0,99176

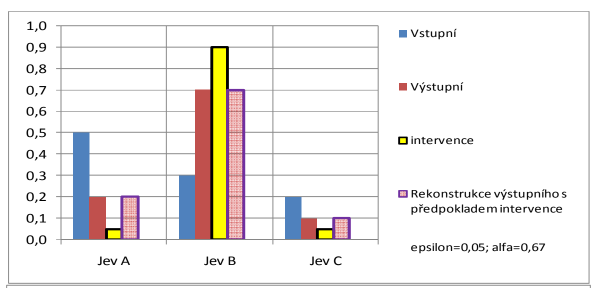


epsilon=	0,50000
alfa=	0,01536

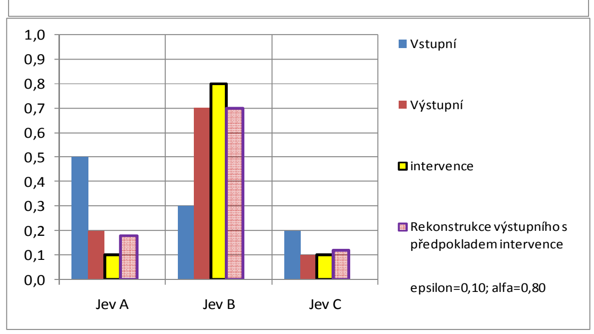
Prezentace jak je možné aproximovat (zde přesně) k danému (libovolnému)spikleneckému intervenčnímu rozdělení jeho sílu.

Další příklad takové rekonstrukce:

	Vstupní	*Výstupní*	*intervence*	Rekonstrukce výstupního s předpokladem intervence
Jev A	0,50	0,20	0,05	0,20
Jev B	0,30	0,70	0,90	0,70
Jev C	0,20	0,10	0,05	0,10



	Vstupní	*Výstupní*	*intervence*	Rekonstrukce výstupního s předpokladem intervence
Jev A	0,50	0,20	0,10	0,18
Jev B	0,30	0,70	0,80	0,70
Jev C	0,20	0,10	0,10	0,12

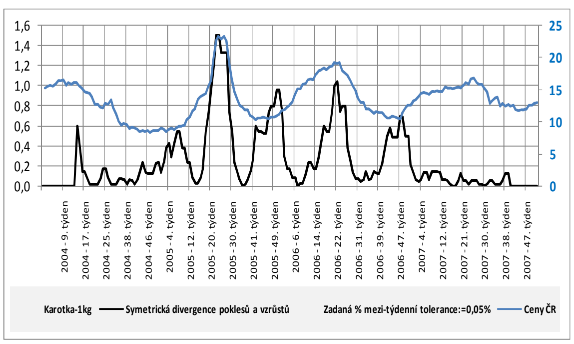


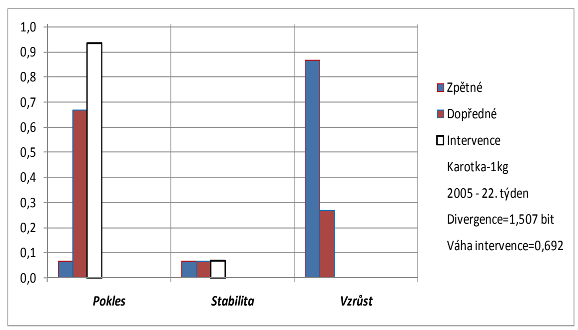
Prezentace jak je možné aproximovat (zde přibližně) k danému (libovolnému) spikleneckému intervenčnímu rozdělení jeho sílu.

Příklad opačného postupu, hledání intervence, k cenovým průběhům (např. sezónní podněty):

Průměrné týdenní ceny vybraných komodit za celou ČR z šetření ČSÚ
Zavedené stavy této časové řady:
$\varepsilon = 0{,}0005$
Stavová diferenční paměť (doba, ze které jsou odhadovány pravděpodobnosti výskytu ve stavu) 12 týdnů zpět a dopředu včetně aktuálního týdne.

Jev	Číselný kód = $dif(t)$	Podmínka
Pokles	-1	$\frac{c(t)-c(t-1)}{c(t)} \leq -\varepsilon$
Stabilita	0	$\frac{\lvert c(t)-c(t-1)\rvert}{c(t)} < \varepsilon$
Růst	+1	$\frac{c(t)-c(t-1)}{c(t)} \geq +\varepsilon$



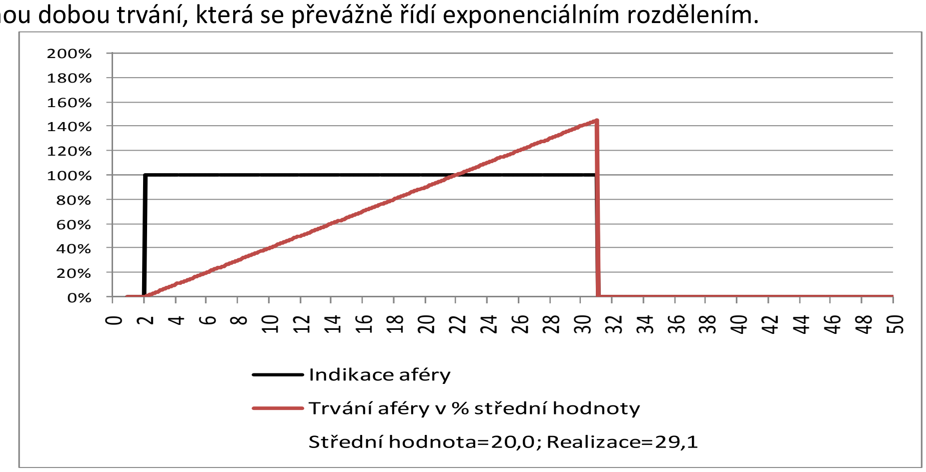


Situace v rozděleních stavového popisu v maximu divergence.

Dynamika „afér“

Spouštění „afér“

Každá „aféra“ je spuštěna nějakým podnětem. Technickým modelem pro ni je „jednopulzní spoušť“, s náhodnou dobou trvání, která se převážně řídí exponenciálním rozdělením.



„Prodlužování afér“

Pokud přijde v době trvání (téže) aféry další podnět, pak se je generována doba odezvy na základě již naběhlé délky reakce na **nejbližší předcházející podnět** a to tak, že **střední hodnota trvání reakce** odpovídající tomuto podnětu je rovna:

$$\textit{střední hodnota reakce na předcházející podnět}\,\frac{\textit{střední hodnota reakce na předcházející podnět}}{\textit{doba trvání reakce na předcházející podnět}} =$$

$$= \frac{(\textit{střední hodnota reakce na předcházející podnět})^2}{\textit{doba trvání reakce na předcházející podnět}}$$

Tedy doby odezev na podněty mají jakousi „martingalovou“ vlastnost. Střední doba reakce následujícího podnětu je dána realizací (náhodou) podnětu bezprostředně předcházejícího.
To umožňuje modelovat procesy „rozvinutí“ aféry a i její doznívání.
Pokud reakce na předchozí podnět již odezněla, je spuštěna nezávisle aféra nová. Pracujeme tedy s dvěma náhodnými proměnnými:

1. Příznak trvání aféry.
2. Doba odezvy na podnět spouštějící aféru.

Metodika má jeden podstatný předpoklad ve dvou rovinách:

1. Teoretická rovina – dva podněty spouštějící aféru nemohou přijít současně.
2. Praktická rovina – (maskování) efekt dalšího podnětu nemá vliv, pokud přijde dostatečně brzo po předchozím. Tedy všechny podněty, které přijdou v době $t + \tau$, kde t je čas posledního podnětu, který nebyl vymaskován, jsou maskovány. Za poslední předchozí podnět je pak uvažován poslední předcházející podnět, který nebyl vymaskován.

Praktický příklad

Aférologický pohled na výskyty článků s tématikou Západočeské univerzity ve vazbě na její právnickou fakultu.

Jednotkou je jeden článek na serveru IDNES, časová jednotka sledování je jeden rok. Byly užity četnosti z let 2005-2010.

Vyhledávání bylo realizováno pomocí množin klíčových slov, vyhledávač neumožňoval využití klíčových sekvencí. Vyhledávač respektoval „ohýbání“ českých slov. Adresa vyhledávače je: http://hledej.idnes.cz/.

Zjištěné četnosti jsou v následujících tabulkách:

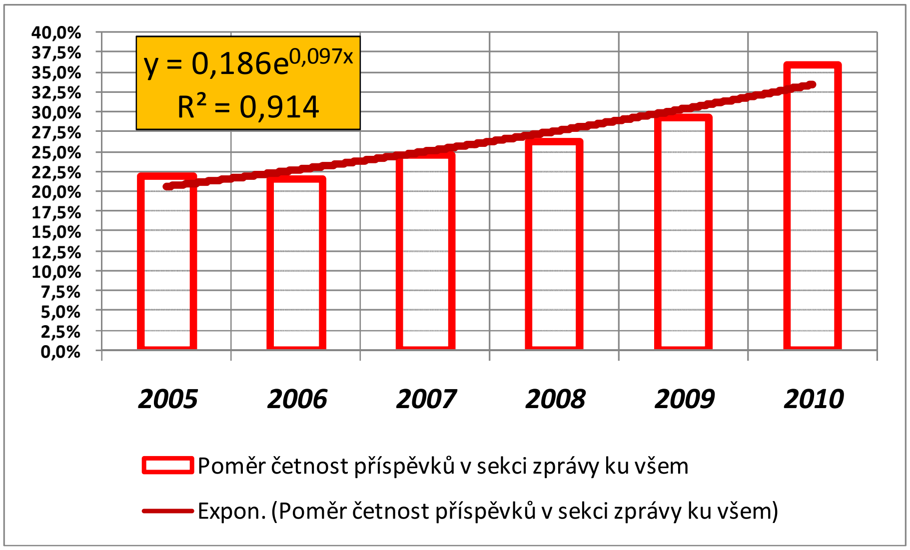
IDNES.CZ - Rubrika: Zprávy, vpravo klíčová slova		Plzeňská právnická fakulta	Západočeská univerzita	fakulta aplikovaných věd	že – jako model libovolného článku
Od	Do	Počet	Počet	Počet	Počet
1.1.2005	31.12.2005	1	4	0	8 427
1.1.2006	31.12.2006	1	13	2	8 290
1.1.2007	31.12.2007	1	13	3	10 453
1.1.2008	31.12.2008	7	29	4	12 360
1.1.2009	31.12.2009	119	87	9	14 447
1.1.2010	31.12.2010	84	84	3	18 288
IDNES.CZ - všechny rubriky, vpravo klíčová slova		Plzeňská právnická fakulta	Západočeská univerzita	fakulta aplikovaných věd	že – jako model libovolného článku
Od	Do	Počet	Počet	Počet	Počet
1.1.2005	31.12.2005	3	16	0	38 335
1.1.2006	31.12.2006	1	20	3	38 373
1.1.2007	31.12.2007	2	23	7	42 570
1.1.2008	31.12.2008	12	42	8	47 176
1.1.2009	31.12.2009	125	103	10	49 495
1.1.2010	31.12.2010	92	93	6	51 135

Aférologická (dojmologická) charakterizace uvedeného období vztažená k ZČU:

IDNES.CZ - všechny rubriky		Typická aféra nebo téma (nebo dezinformace) spojené se sledovaným tématem a obdobím (citace).
Od	Do	
1.1.2005	31.12.2005	***Gross by měl vydat svou rigorózní práci, soudí děkan.***
1.1.2006	31.12.2006	***František Mrázek měl úzké vztahy s policisty. Teď je vyšetřuje Inspekce ministra vnitra. Objevila u něj studijní doklad mladého policisty z Odboru pro odhalování závažné obecné kriminality.***
1.1.2007	31.12.2007	***Studentka Západočeské univerzity konkuruje pornoherečkám.***
1.1.2008	31.12.2008	***Jeden připravila společnost Consulting Company Novasoft spolu s firmou SpeechTech zřízenou Západočeskou univerzitou v Plzni, druhý Technická univerzita v Liberci spolupracující s firmou Newton IT.***
1.1.2009	31.12.2009	***Proděkan na plzeňských právech čelí obvinění z plagiátorství.***
1.1.2010	31.12.2010	***O etice na plzeňské univerzitě budou rozhodovat její vlastní akademici***

Empirické pozorování

Vývoj „čistých zpráv“ na úkor ne-zpráv (komentářů a volných textů):



Střední meziroční poměr 1.102.

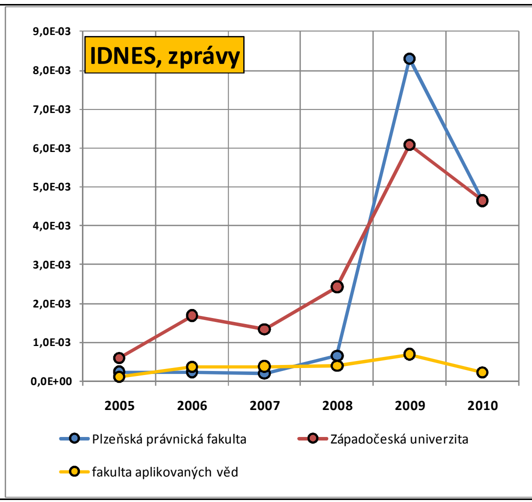
Problém současných médií je v tom, že relativně málo sdělují zprávy na úkor zpráv o zprávách, tedy komentáře, názory a autorské texty. Tedy dostáváme informaci nikoliv o tom, co se stalo, ale co si o tom, co se stalo, někdo myslí, nebo co si o tom máme myslet. Leč v tomto konkrétním případě lze pozorovat mírné zlepšování ale dle vlastní klasifikace daného média.

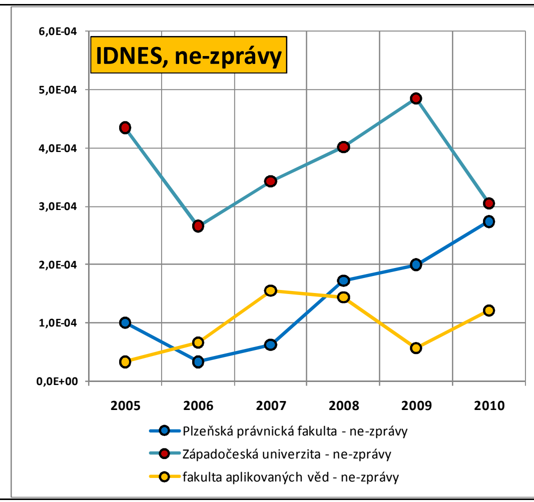
Vstupní data:

Rok a počty všech článků	IDNES.CZ - Rubrika: Zprávy	IDNES.CZ - všechny rubriky	Poměr četnost příspěvků v sekci zprávy ke všem
	Počet	Počet	%
2005	8 427	38 335	22,0%
2006	8 290	38 373	21,6%
2007	10 453	42 570	24,6%
2008	12 360	47 176	26,2%
2009	14 447	49 495	29,2%
2010	18 288	51 135	35,8%

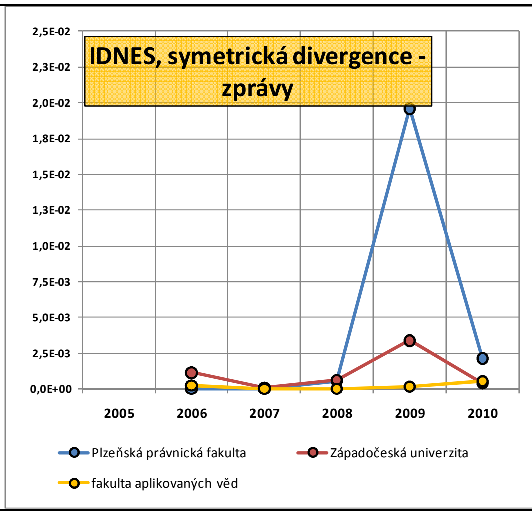
Výsledky

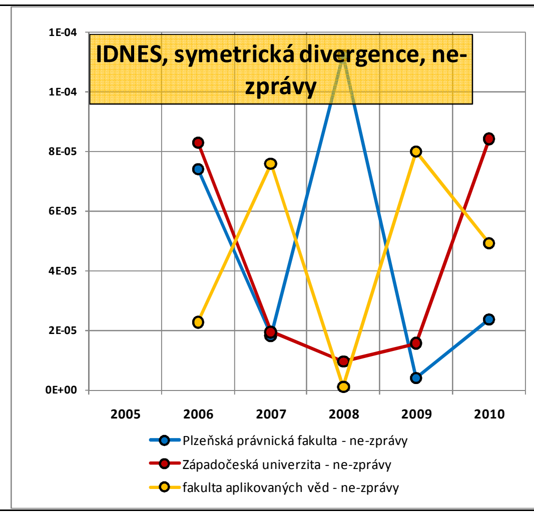
Dynamika





Grafy odhadů pravděpodobností výskytu „příspěvku“ s uvedenými klíčovými slovy.





Grafy odhadů divergencí meziročních pravděpodobnostních rozdělení alternativního jevu výskytu „příspěvku“ s uvedenými klíčovými slovy. Použit logaritmus se základem ***e***. Jeffreysova J-divergence.

Prameny:

[S]	SHANNON, C. E.	A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
[1]	NEUMAN, F.	Funkcionální rovnice, SNTL, Praha 1986
[2]	RÉNYI, A.	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
[3]	RAO, C. R.	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
[C]	Thomas M. Cover, Joy A. Thomas	Elements of Information Theory. 1991 John Wiley & Sons, Inc. ISBN 0-471-06259-6.

Příloha

Podmíněná nezávislost neimplikuje nezávislost

$$P(x, y/z) = P(x/z)P(y/z)$$

P(x,y/z=0)

x,y->	0	1	p(x/z=0)
0	0,140	0,560	0,700
1	0,060	0,240	0,300
p(y/z=0)	0,200	0,800	1,000

p(z=0)=	**0,250**
p(z=1)=	**0,750**

P(x,y/z=1)

x,y->	0	1	p(x/z=1)
0	0,200	0,200	0,400
1	0,300	0,300	0,600
p(y/z=1)	0,500	0,500	1,000

P(x,y,z=0)

x,y->	0	1	p(x,z=0)
0	0,035	0,140	0,175
1	0,015	0,060	0,075
p(y,z=0)	0,050	0,200	0,250

P(x,y,z=1)

x,y->	0	1	p(x,z=1)
0	0,150	0,150	0,300
1	0,225	0,225	0,450
p(y,z=1)	0,375	0,375	0,750

P(x,y)

x,y->	**0**	**1**	**p(x)**

0	0,185	0,290	0,475
1	0,240	0,285	0,525
p(y)	0,425	0,575	1,000

P(x)P(y)

x,y->	0	1	p(x)
0	0,202	0,273	0,475
1	0,223	0,302	0,525
p(y)	0,425	0,575	1,000

Zde je $P(x,y) \neq P(x)P(y)$

Nezávislost neimplikuje podmíněnou nezávislost

$$P(x,y) = P(x)P(y)$$

P(x,y/z=0)

x,y->	0	1	p(x/z=0)
0	0,096	0,512	0,608
1	0,096	0,296	0,392
p(y/z=0)	0,192	0,808	1,000

P(x/z=0)P(y/z=0)

x,y->	0	1	p(x/z=0)
0	0,117	0,491	0,608
1	0,075	0,317	0,392
p(y/z=0)	0,192	0,808	1,000

p(z=0)=	0,250
p(z=1)=	0,750

P(x,y/z=1)

x,y->	0	1	p(x/z=1)
0	0,075	0,256	0,331
1	0,128	0,541	0,669
p(y/z=1)	0,203	0,797	1,000

P(x/z=1)P(y/z=1)

x,y->	0	1	p(x/z=1)
0	0,067	0,264	0,331
1	0,136	0,534	0,669
p(y/z=1)	0,203	0,797	1,000

P(x,y,z=0)

x,y->	0	1	p(x,z=0)
0	0,024	0,128	0,152
1	0,024	0,074	0,098
p(y,z=0)	0,048	0,202	0,250

P(x,y,z=1)

x,y->	0	1	p(x,z=1)
0	0,056	0,192	0,248
1	0,096	0,406	0,502
p(y,z=1)	0,152	0,598	0,750

P(x,y)

x,y->	0	1	p(x)
0	0,080	0,320	0,400

1	0,120	0,480	0,600
p(y)	0,200	0,800	1,000

P(x)P(y)

x,y->	0	1	p(x)
0	0,080	0,320	0,400
1	0,120	0,480	0,600
p(y)	0,200	0,800	1,000

Zde je $P(x, y / z) \neq P(x / z) P(y / z)$