FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

TEORIE INFORMACE
STRUČNÝ ÚVOD PRO KURZ MSTU

Autor textu:
Ing. Petr Honzík, Ph.D.

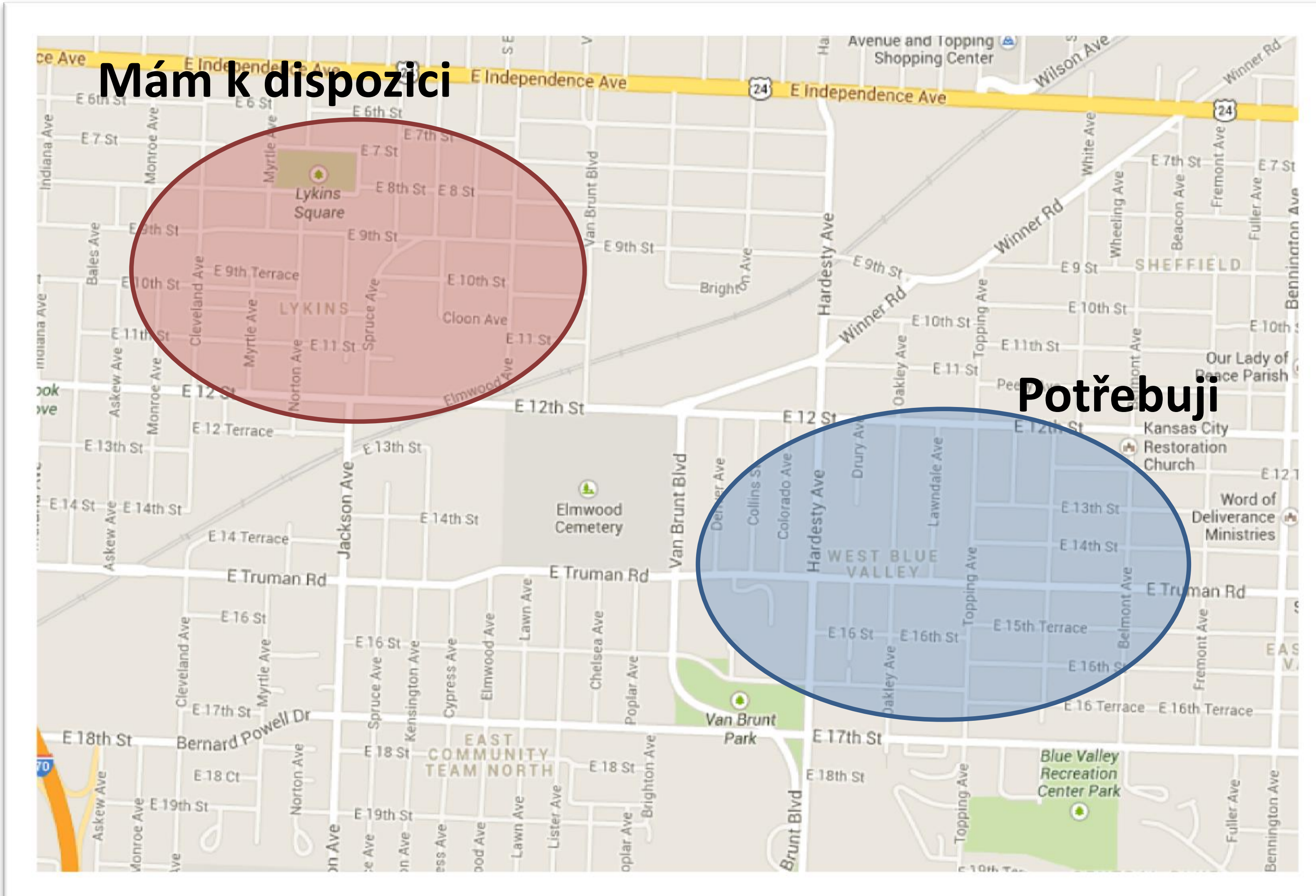
Květen 2014

Komplexní inovace studijních programů a zvyšování kvality výuky na FEKT VUT v Brně
OP VK CZ.1.07/2.2.00/28.0193

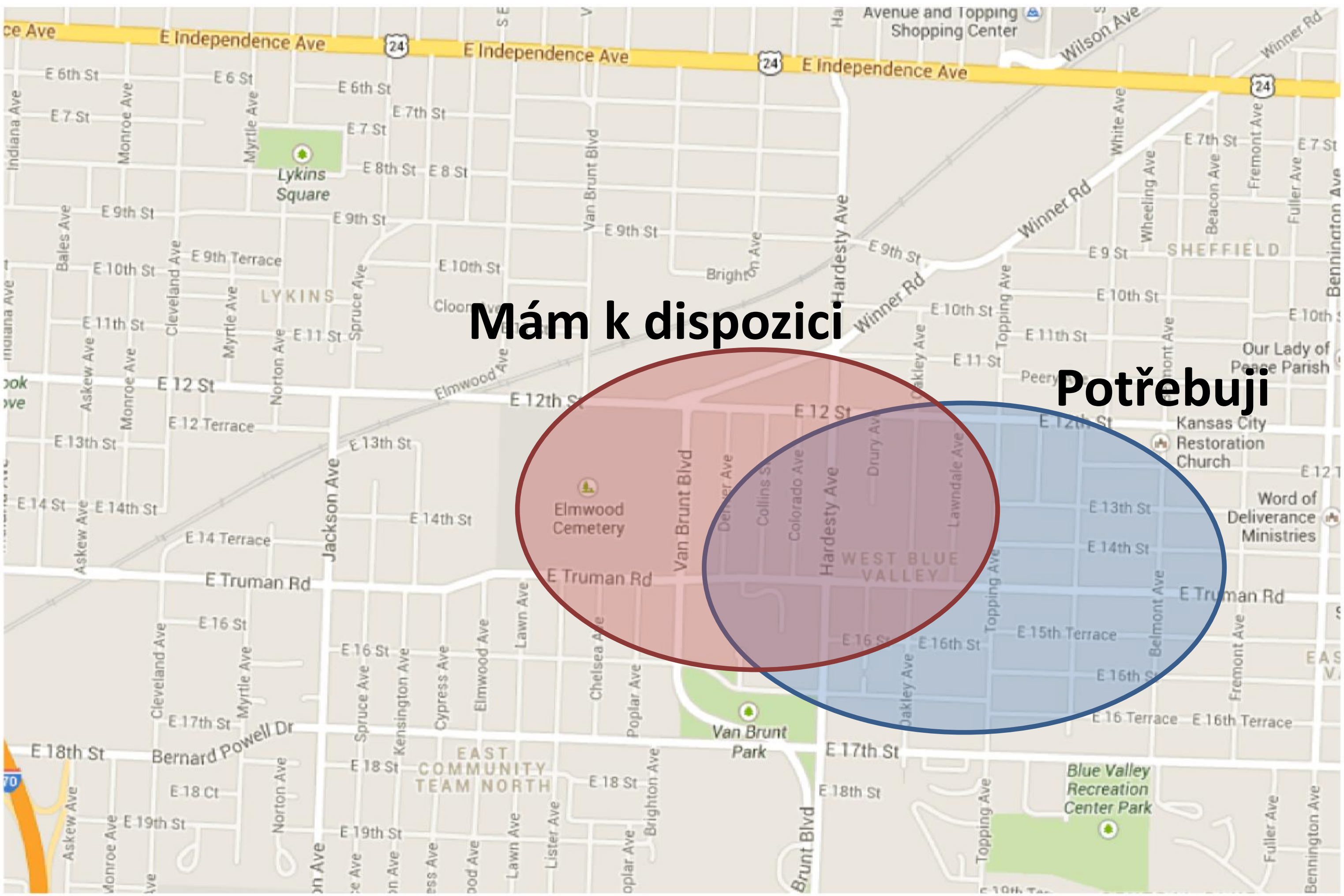
evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

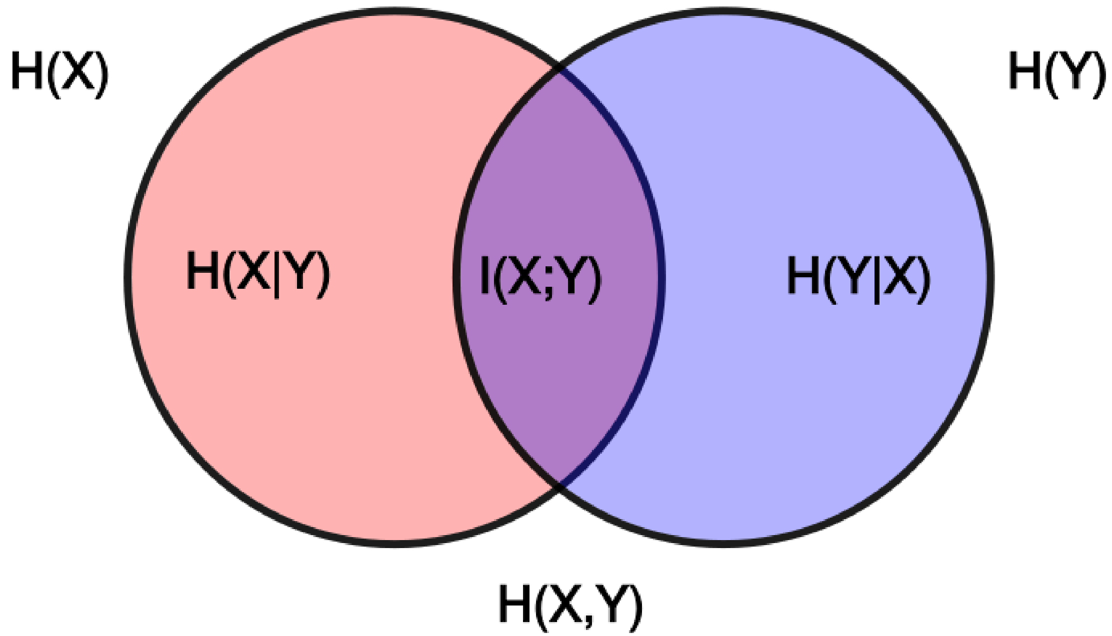
TI – intuitivně



TI – intuitivně

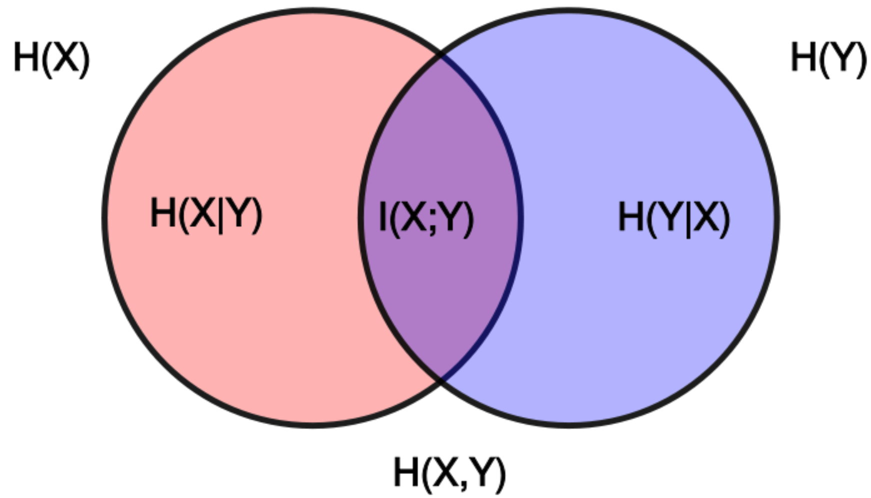


TI – základní pojmy



H(X)	– entropie (*entropy, marginal entropy*)
H(X,Y)	– sdružená entropie (*joint entropy*)
H(X|Y)	– podmíněná entropie (*conditional entropy*)
I(X,Y)	– vzájemná informace (*mutual information*)

TI – základní pojmy



Odvoďte:

I(X,Y) = H(X,Y) … ?
H(Y|X) = H(Y) … ?
H(X)+H(Y) = H(X,Y) …?
I(X,X) = … ?

Možná řešení (může jich být více)

I(X,Y) = H(X,Y) - H(X|Y) - H(Y|X)
H(Y|X) = H(Y) – I(X,Y)
H(X)+H(Y) = H(X,Y) + I(X,Y)
I(X,X) = H(X) (*self-information*)

Kvantifikace informace

- Je dána množina hodnot $\mathcal{X}$. Množství informace získané zjištěním konkrétní hodnoty z množiny $\mathcal{X}$ je definováno jako **záporný logaritmus pravděpodobnosti výběru** této hodnoty.

$$I(x) = -\log p(x)$$

- Automaticky je předpokládáno, že jednotlivé výběry jsou náhodné nezávislé jevy.
- Logaritmus bude nadále počítán se základem 2, jednotkou je [bit] (základ 2 však není podmínkou).

Kvantifikace informace – příklady

- Je-li tedy $\mathcal{X} = \{a, b, c\}$, množství informace získané zjištěním hodnoty a:

$$I(a) = -\log p(a) = -\log \frac{1}{3} = 1{,}58$$
$$I(a) = I(b) = I(c)$$

- Je-li šest různých měření hodnoty z množiny $\mathcal{X}$ definovánu vektorem $\boldsymbol{x} = [a, b, b, a, c, a]$

$$I(a) = -\log p(a) = -\log \frac{3}{6} = 1$$
$$I(b) = -\log p(b) = -\log \frac{2}{6} = 1{,}58$$
$$I(c) = -\log p(c) = -\log \frac{1}{6} = 2{,}58$$

Entropie

- **střední hodnota informace** $I(X)$ získaná zjištěním konkrétní hodnoty X

$$H(X) = E[I(X)] = -\sum_{\forall x \in \boldsymbol{\mathcal{X}}} p(x) I(x) = -\sum_{\forall x \in \boldsymbol{\mathcal{X}}} p(x) \log p(x)$$

- Je–li $\boldsymbol{x} = [a, b, b, a, c, a]$, pak

$$H(\boldsymbol{x}) = E[I(\boldsymbol{x})] = -\sum_{\forall x \in \boldsymbol{\mathcal{X}}} p(x) \log p(x) =$$

$$-\frac{3}{6}\log\frac{3}{6} - \frac{2}{6}\log\frac{2}{6} - \frac{1}{6}\log\frac{1}{6} = \frac{3}{6}1 + \frac{2}{6}1{,}58 + \frac{1}{6}2{,}58 = 1{,}46$$

Entropie ze sdružené pravděpodobnosti

- *p*(x) je tzv. **marginální pravděpodobnost** ve sdružené pravděpodobnosti a je definována:

$$p(x) = \sum_{\forall y \in Y} p(x, y)$$

- entropii veličiny X pak lze zapsat jako:

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

Sdružená a podmíněná entropie

- sdružená entropie je definována přímo z matice sdružené pravděpodobnosti:

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

- podmíněná pravděpodobnost je definována jako:

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

- z čehož lze dosazením získat podmíněnou entropii:

$$H(X|Y) = -\sum_{\forall x \in X} \sum_{\forall y \in Y} p(x,y) \log p(x|y) = -\sum_{\forall x \in X} \sum_{\forall y \in Y} p(x,y) \log \frac{p(x,y)}{p(y)}$$

Vzájemná informace

- vzájemná informace

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- je odvozena z kombinací předešlého jako:

$$-\sum_{\forall x \in X}\sum_{\forall y \in Y} p(x,y)\log p(x) + \sum_{\forall x \in X}\sum_{\forall y \in Y} p(x,y)\log\frac{p(x,y)}{p(y)} =$$

$$\sum_{\forall x \in X}\sum_{\forall y \in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

- Poznámka: jsou-li veličiny *X* a *Y* nezávislé, potom platí, že $p(x,y)=p(x)p(y)$. Podíl těchto pravděpodobností je roven 1 a log1 = 0. Tedy pro dvě nezávislé veličiny bude vzájemná informace rovna 0.

Informační zisk (IG)

- **informační zisk** (information gain) je synonymum pro
  - vzájemnou informaci (tak bude chápáno v kurzu MSTU; použití v rozhodovacích stromech, výběru příznaků)
  - Kullback-Leiblerovu divergenci, množství chybějící informace, je-li použito optimální kódování dle rozložení *q*, avšak skutečné je *p*

$$D_{KL}(p,q) = \sum_{\forall y \in Y} p \log \frac{p}{q}$$

  shodnost s *IG* však platí pouze pokud D_{KL} je definováno jako funkce veličin *X* a *Y;* je-li D_{KL} definováno jako funkce rozložení *p* a *q*, tak pro něj neplatí axiom symetrie

- existuje více způsobů normalizace IG (*information gain ratio, symmetric uncertainty, normalized information gain*)