Teorie kódování - Kapitola 3
Komunikace kanály se šumem

Jan Paseka

Ústav matematiky a statistiky
Masarykova univerzita

26. dubna 2023

Obsah

Základní pojmy

FI Komunikační systém je mechamismus, který zprostředkovává přenos informace od zdroje zprávy až k zařízení, které tuto informaci zpracovává. Obecně sestává z kodéru, sdělovacího (komunikačního) kanálu a následně dekodéru.

Základní pojmy

FI Komunikační systém je mechamismus, který zprostředkovává přenos informace od zdroje zprávy až k zařízení, které tuto informaci zpracovává. Obecně sestává z kodéru, sdělovacího (komunikačního) kanálu a následně dekodéru.

Obrázek 1: Blokový diagram obecného sdělovacího systému.

Základní pojmy

FI Komunikační systém je mechamismus, který zprostředkovává přenos informace od zdroje zprávy až k zařízení, které tuto informaci zpracovává. Obecně sestává z kodéru, sdělovacího (komunikačního) kanálu a následně dekodéru.

Obrázek 1: Blokový diagram obecného sdělovacího systému.

Zdrojová zpráva je obvykle v podobě sekvence binárních nebo desítkových číslic, nebo sekvence abecedních znaků převedených do technicky zpracovatelné podoby.

Základní pojmy

Kódovací zařízení převádí tyto zprávy na signály kompatibilní se vstupem kanálu – obvykle se jedná o elektrické signály, které mají jistá omezení na velikost napětí, šířku pásma a délku trvání impulzu.

Základní pojmy

Kódovací zařízení převádí tyto zprávy na signály kompatibilní se vstupem kanálu – obvykle se jedná o elektrické signály, které mají jistá omezení na velikost napětí, šířku pásma a délku trvání impulzu.

Takto upravené pak vstupují do sdělovacího kanálu a jsou vystaveny šumu (tj. možnosti vzniku chyby). Výstup z kanálu vstupuje do dekodéru, jehož funkcí je rozhodnout, jakou podobu měla původní zdrojová zpráva, a tu pak přivést na výstup celého sdělovacího systému.

Základní pojmy

Většina komunikačních kanálů má konečnou kapacitu přenosu informace, tj. míru schopnosti přenášet informaci měřenou v bitech za sekundu nebo bitech na symbol.

Základní pojmy

Většina komunikačních kanálů má konečnou kapacitu přenosu informace, tj. míru schopnosti přenášet informaci měřenou v bitech za sekundu nebo bitech na symbol.

Díky vynikající teoretické práci Shannona (r. 1948) se dá ukázat, že pokud je průměrná rychlost přenosu informace menší než kapacita kanálu, je možné vybrat množinu signálů (kódových slov) takovou, že pravděpodobnost výskytu chyby při dekódování bude libovolně malá.

Základní pojmy

Většina komunikačních kanálů má konečnou kapacitu přenosu informace, tj. míru schopnosti přenášet informaci měřenou v bitech za sekundu nebo bitech na symbol.

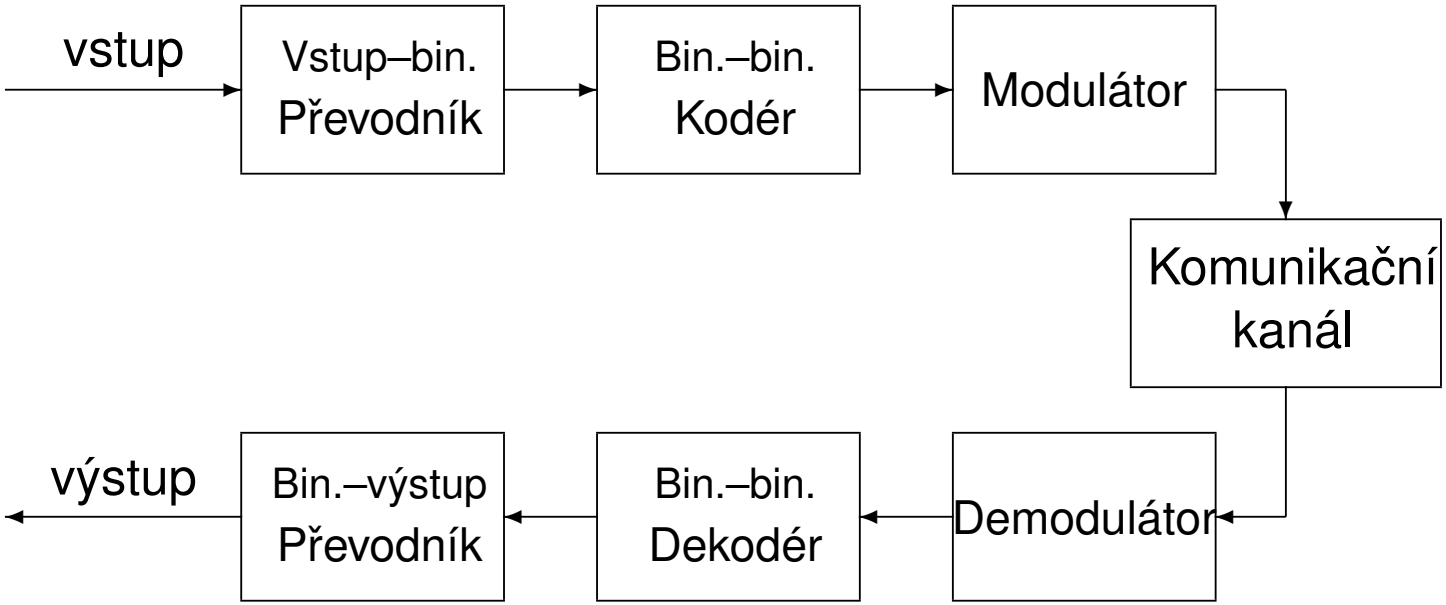
Díky vynikající teoretické práci Shannona (r. 1948) se dá ukázat, že pokud je průměrná rychlost přenosu informace menší než kapacita kanálu, je možné vybrat množinu signálů (kódových slov) takovou, že pravděpodobnost výskytu chyby při dekódování bude libovolně malá.

Nicméně jakkoliv je tato teorie mocná, výsledek nevypovídá nic o tom, jak tyto signály volit, ani zda je možné je pomocí současných technických prostředků konstruovat.

Používání samoopravných kódů je pokusem výše uvedené dva problémy obejít.

Základní pojmy

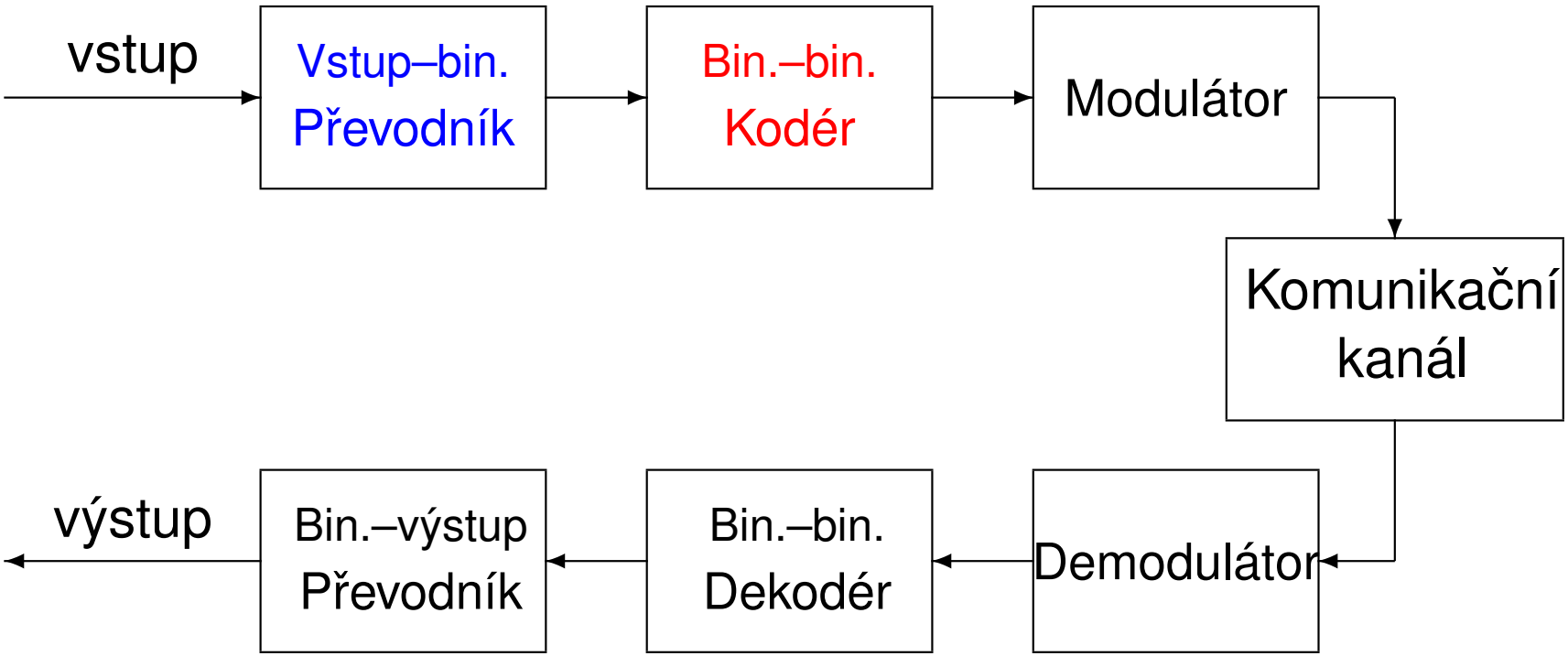
Ovšem celý přenos informace od zdroje až po zpracování není tak jednoduchý, jak znázorňuje (obr.4). Sestává z komplexnějšího sdělovacího systému; jednu takovou možnost nabízí (obr.2).



Obrázek 2: Konkrétní sdělovací systém.

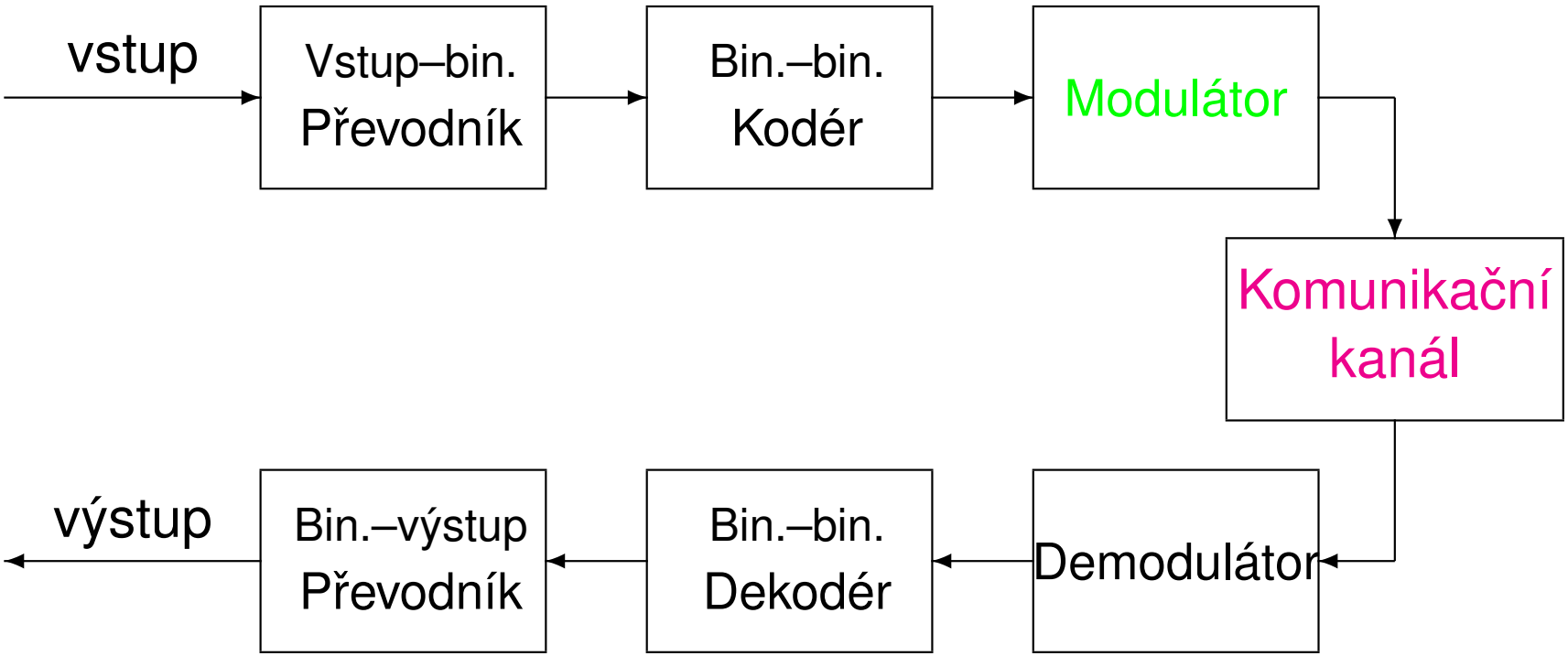
Základní pojmy

Kodéry (převodníky) převádí znaky jedné abecedy na znaky abecedy jiné. Obvykle mají obě abecedy poměrně malou mohutnost – typický převod může být z desítkové do dvojkové soustavy.



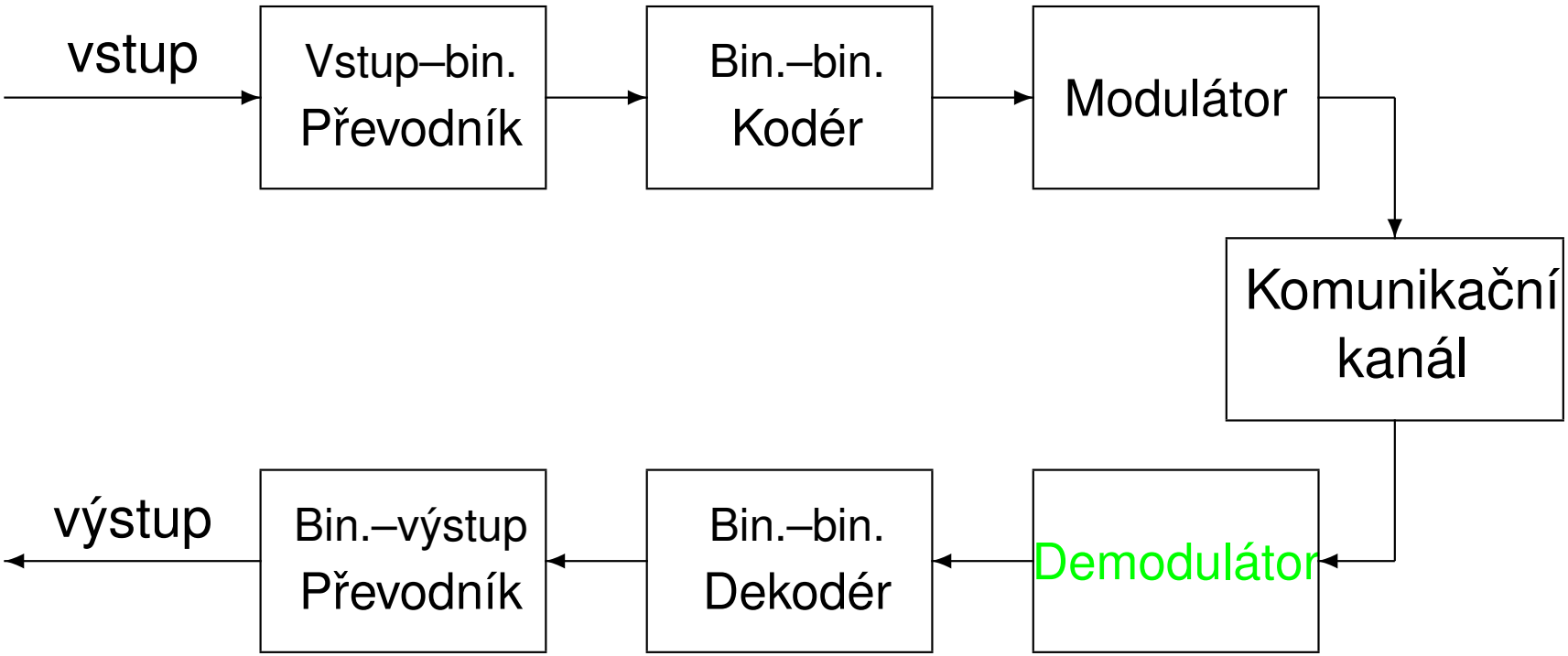
Základní pojmy

Modulátor na vstupu přijímá jednotlivé znaky a ke každému znaku vytváří proudový impuls, který vstupuje do kanálu. Tato operace s každým znakem zvlášť je omezením při přenosu informace a způsobí tak ztrátu kapacity kanálu.



Základní pojmy

Demodulátor provádí inverzní operaci. Ke každému obdrženému impulsu hledá znak tak, aby pravděpodobnost přenosové chyby byla co nejmenší. A opět jako při modulaci, i zde individuální modulace způsobuje ztrátu kapacity.



Obsah

Základní pojmy

FI Ve svém nejširším smyslu lze komunikační kanál považovat za černou skříňku, která akceptuje řetězce symbolů ze vstupní abecedy Σ_1 a vysílá řetězce symbolů z výstupní abecedy Σ_2.

Základní pojmy

FI Ve svém nejširším smyslu lze komunikační kanál považovat za černou skříňku, která akceptuje řetězce symbolů ze vstupní abecedy Σ_1 a vysílá řetězce symbolů z výstupní abecedy Σ_2.

Můžeme zřejmě tvrdit jen málo o takovéto struktuře. Omezme pozornost na ***diskrétní kanál bez paměti***, který je charakterizován **vstupní abecedou** $\Sigma_1 = \{a_1, \ldots, a_m\}$, **výstupní abecedou** $\Sigma_2 = \{b_1, \ldots, b_n\}$ a ***maticí P kanálu***

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & \cdots & p_{1n-1} & p_{1n} \\ p_{21} & p_{22} & \cdots & \cdots & p_{2n-1} & p_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots & \vdots \\ p_{m-11} & p_{m-12} & \cdots & \cdots & p_{m-1n-1} & p_{m-1n} \\ p_{m1} & p_{m2} & \cdots & \cdots & p_{mn-1} & p_{mn} \end{pmatrix}.$$

Základní pojmy

Způsob používání kanálu je následující: každá posloupnost $(u_1, u_2, \ldots, u_N)$ symbolů ze vstupní abecedy Σ_1 na vstupu se převede na posloupnost $(v_1, v_2, \ldots, v_N)$ téže délky symbolů z výstupní abecedy Σ_2 na výstup tak, že

$$P(v_k = b_j | u_k = a_i) = p_{ij} \quad (1 \leq i \leq m, 1 \leq j \leq n),$$

a to nezávisle pro každé k.

Základní pojmy

Způsob používání kanálu je následující: každá posloupnost $(u_1, u_2, \ldots, u_N)$ symbolů ze vstupní abecedy Σ_1 na vstupu se převede na posloupnost $(v_1, v_2, \ldots, v_N)$ téže délky symbolů z výstupní abecedy Σ_2 na výstup tak, že

$$P(v_k = b_j | u_k = a_i) = p_{ij} \quad (1 \leq i \leq m, 1 \leq j \leq n),$$

a to nezávisle pro každé k.

Implicitně je ve výše uvedeném obsaženo, že pro každé i, $1 \leq i \leq m$ platí

$$\sum_j p_{ij} = 1.$$

Základní pojmy

Matice P s nezápornými hodnotami taková, že součet prvků v každém řádku je roven 1, se nazývá ***stochastická matice***.

Základní pojmy

Matice P s nezápornými hodnotami taková, že součet prvků v každém řádku je roven 1, se nazývá ***stochastická matice***.

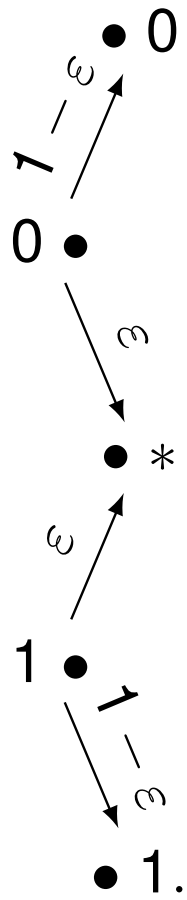
V teorii náhodných procesů mluvíme o matici přechodu markovského řetězce.

Základní pojmy

Matice P s nezápornými hodnotami taková, že součet prvků v každém řádku je roven 1, se nazývá ***stochastická matice***.

V teorii náhodných procesů mluvíme o matici přechodu markovského řetězce.

Je často užitečné reprezentovat kanál pomocí diagramu, jako je tomu např v následujícím Příkladu 2.1.



Příklady

Příklad 2.1

Binární vypouštěcí kanál *má vstupní abecedu* $\Sigma_1 = \{0, 1\}$, *výstupní abecedou* $\Sigma_2 = \{0, 1, *\}$ *a* maticí P kanálu

$$P = \begin{pmatrix} 1-\varepsilon & 0 & \varepsilon \\ 0 & 1-\varepsilon & \varepsilon \end{pmatrix}.$$

To odpovídá situaci, pro kterou má každý symbol pravděpodobnost ε*, že se špatně přenese a to na* $*$*. Ale jak 1 tak 0 nelze navzájem zaměnit.*

Příklady

Příklad 2.2

Nejpoužívanější kanál v tomto modelu komunikace **binární symetrický kanál** *má vstupní abecedu* $\Sigma_1 = \{0, 1\}$*, výstupní abecedou* $\Sigma_1 = \{0, 1\}$ *a* maticí P kanálu

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

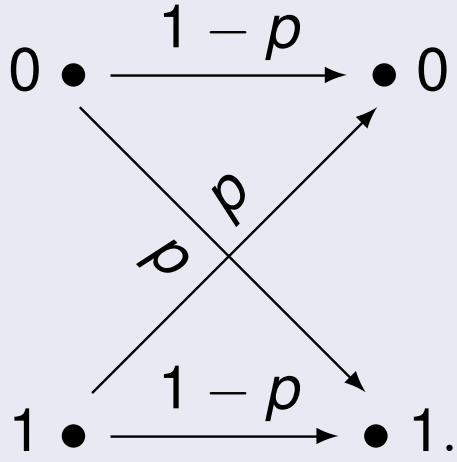
Příklady

Příklad 2.2

Nejpoužívanější kanál v tomto modelu komunikace **binární symetrický kanál** *má vstupní abecedu* $\Sigma_1 = \{0, 1\}$, *výstupní abecedou* $\Sigma_1 = \{0, 1\}$ *a* maticí P kanálu

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

Diagram odpovídající tomuto kanálu má tvar



Jinak řečeno, to odpovídá situaci, pro kterou má každý symbol x pravděpodobnost p, že se špatně přenese a to na $1 - x$. *Často budeme psát* $q = 1 - p$ *bez dalšího komentáře.*

Rozšíření diskrétních kanálů bez paměti

Uvažme diskrétní kanál bez paměti se vstupní abecedu Σ_1, výstupní abecedou Σ_2 a *maticí P kanálu* . *r-té rozšíření* tohoto kanálu je diskrétní kanál bez paměti se vstupní abecedu $\Sigma_1^{(r)}$, výstupní abecedou $\Sigma_2^{(r)}$ a *maticí* $P^{(r)}$ *kanálu* , která je definována následovně:
Souřadnice (i, j) matice $P^{(r)}$ odpovídající vstupu

$$\sigma_i = \alpha_1 \alpha_2 \dots \alpha_r$$

s $\alpha_k \in \Sigma_1$, a výstupu

$$\tau_j = \beta_1 \beta_2 \dots \beta_r$$

s $\beta_k \in \Sigma_2$, je

$$(P^{(r)})_{ij} = p(\beta_1|\alpha_1)p(\beta_2|\alpha_2)\dots p(\beta_r|\alpha_r),$$

Rozšíření diskrétních kanálů bez paměti

Souřadnice (i, j) matice $P^{(r)}$ odpovídající vstupu

$$\sigma_i = \alpha_1 \alpha_2 \dots \alpha_r$$

s $\alpha_k \in \Sigma_1$, a výstupu

$$\tau_j = \beta_1 \beta_2 \dots \beta_r$$

s $\beta_k \in \Sigma_2$, je

$$(P^{(r)})_{ij} = p(\beta_1|\alpha_1)p(\beta_2|\alpha_2)\dots p(\beta_r|\alpha_r),$$

kde $p(\beta_k|\alpha_k)$ je pravděpodobnost, že v kanálu s maticí P je obdržen symbol β_k za předpokladu odeslání α_k.

Příklad - Druhé rozšíření

Příklad 2.3

Druhé rozšíření binárního symetrického kanálu *s* maticí *P* kanálu

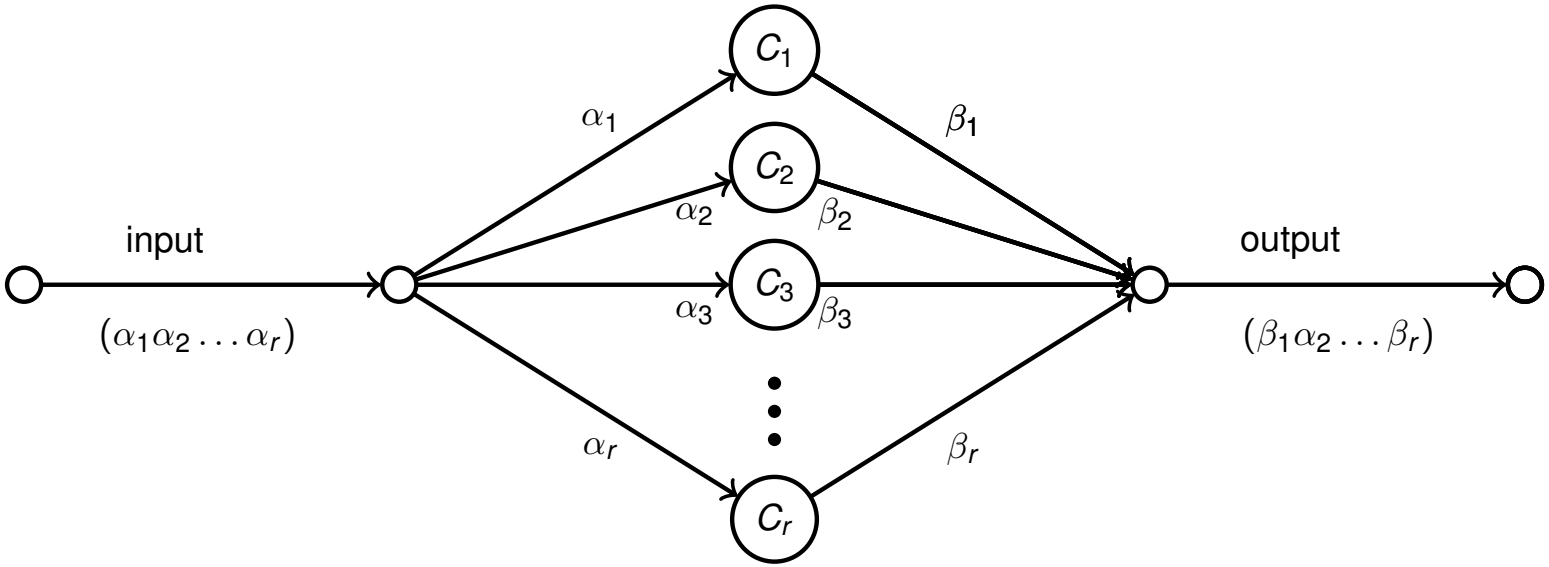
$$P = \begin{pmatrix} q & p \\ p & q \end{pmatrix}$$

má tvar

$$P^2 = \begin{pmatrix} q^2 & qp & pq & p^2 \\ qp & q^2 & p^2 & pq \\ pq & p^2 & q^2 & qp \\ p^2 & pq & qp & q^2 \end{pmatrix} = \begin{pmatrix} qP & pP \\ pP & qP \end{pmatrix}.$$

Rozšíření diskrétních kanálů bez paměti

Alternativní způsob jak můžeme přemýšlet o *r*-tém rozšíření je, že kanál C považujeme za r kopií C operujících nezávisle a paralelně dle níže uvedeného.



Obrázek 3: *r*-té rozšíření

Rozšíření diskrétních kanálů bez paměti - Cvičení

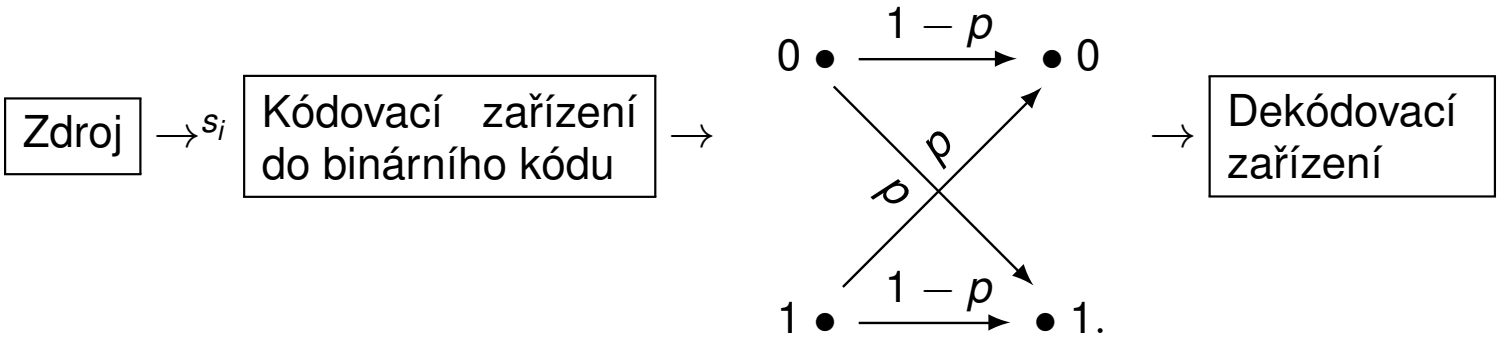
Cvičení 2.4

1. *Zpráva sestávající z N binárních číslic je přenesena binárním symetrickým kanálem mající pravděpodobnost chyby přenosu p. Ukažte, že očekávaný počet chyb je Np.*

Obsah

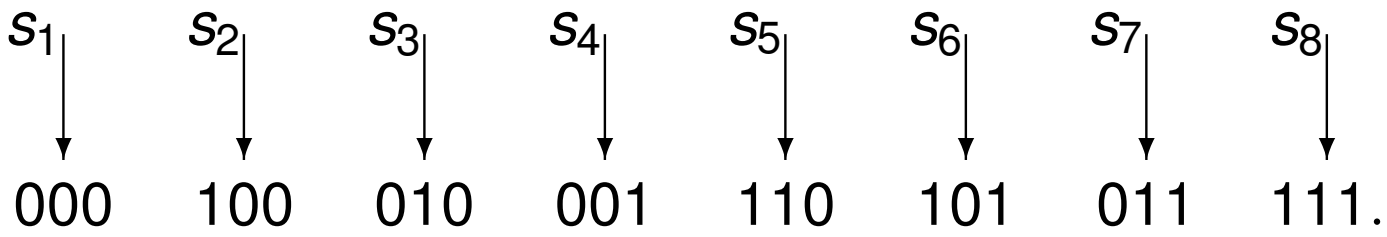
Základní pojmy

FI Uvažme následující situaci: máme zdroj bez paměti $\mathcal{S}$, který vysílá symboly (zdrojová slova) $s_1, \ldots, s_N$ s pravděpodobnostmi $p_1, \ldots, p_N$.
Zdroj je spojen s binárním symetrickým kanálem s pravděpodobností chyby následovně:



Základní pojmy

Budeme předpokládat, že zakódování do binárního kódu proběhne bez šumu a že způsob zakódování je znám dekódovacímu zařízení.
Předpokládejme pro jednoduchost, že $N = 8$; za symboly můžeme považovat např. písmena abecedy, různé měny nebo cokoliv jiného. Efektivní (tj. kompaktní) zakódování ve smyslu předchozího odstavce je pak



Základní pojmy

Zpráva je řetězec zdrojových slov s_i, která jsou postupně zakódovaná, přenesená a pak dekódovaná. Tudíž, dle výše uvedeného kódovacího schématu, je pravděpodobnost, že jisté slovo je správně přeneseno, rovna q^3. Pravděpodobnost, že zpráva o n slovech je korektně přenesena, je q^{3n}.

Základní pojmy

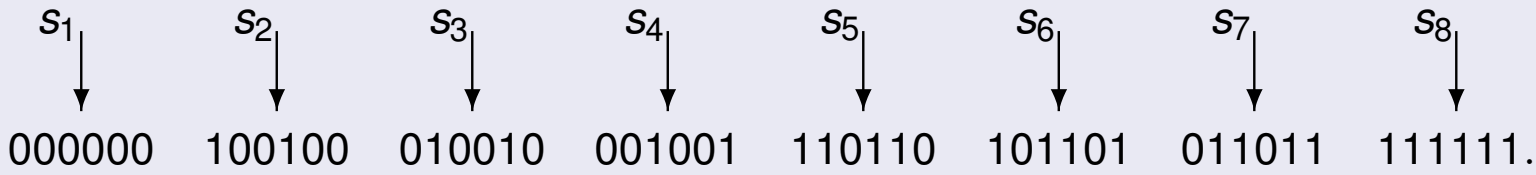
Zpráva je řetězec zdrojových slov s_i, která jsou postupně zakódovaná, přenesená a pak dekódovaná. Tudíž, dle výše uvedeného kódovacího schématu, je pravděpodobnost, že jisté slovo je správně přeneseno, rovna q^3. Pravděpodobnost, že zpráva o n slovech je korektně přenesena, je q^{3n}.

Lze to provést lépe? Odpověď je samozřejmě ano, jinak bychom se vůbec neptali. Zajímavé otázky jsou (a) jak moc lépe a (b) na čí náklady?

Příklady

Příklad 3.1

Uvažme výše uvedený příklad s osmi stejně pravděpodobnými zdrojovými slovy a předpokládejme, že použijeme zdvojené zakódování následovně:



Pokud dekódovací zařízení přijme pravidlo, že bude pouze dekódovat v případě, že první tři symboly a druhé tři symboly jsou totožné, a jinak "zavolá o pomoc", pravděpodobnost, že se vyskytne chyba a zůstane neobjevena, se drasticky redukuje. Zajisté budeme platit podstatnou cenu tím, že jsme snížili poměr přenosu faktorem 2.

Příklady

Příklad 3.1 (Pokračování)

Navíc se jedná o čistě detekční systém, který nebude využitelný v případě, že se dekódovací zařízení nemůže kontaktovat s kódovacím zařízením a požádat ho o znovuposlání slova, u kterého byla detekována chyba.

Příklady

Příklad 3.1 (Pokračování)

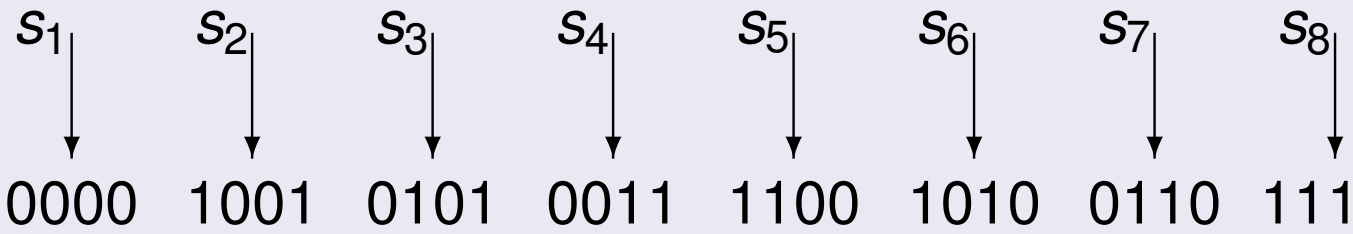
Navíc se jedná o čistě detekční systém, který nebude využitelný v případě, že se dekódovací zařízení nemůže kontaktovat s kódovacím zařízením a požádat ho o znovuposlání slova, u kterého byla detekována chyba.

Zbývající část této kapitoly je věnována způsobu získání dostatečné míry přenosu kanálu se šumem bez příliš velkého prodloužení zprávy.

Cvičení

Cvičení 3.2

Jednoduchý způsob detekování *nejvýše jedné chyby je použít zařízení přidávajícího* kontrolu parity*, abychom měli zajištěno, že součet čísel v přenášeném slově je sudý. Tedy kontrola parity z výše uvedeného příkladu má tvar*



Ukažte, že jestliže přeneseme kód s kontrolou parity binárním symetrickým kanálem, pravděpodobnost, že není objevena chyba, je rovna $6p^2(1-p)^2 + p^4$*, kde p je pravděpodobnost výskytu chyby při přenosu kanálem.*

Obsah

Dekódovací pravidla

FI Buď dán kanál bez paměti se vstupní abecedou $\Sigma_1 = \{a_1, \ldots, a_m\}$ a výstupní abecedou $\Sigma_2 = \{b_1, \ldots, b_k\}$. ***Kód délky*** n je libovolný systém $\mathcal{C}$ různých posloupností délky n symbolů ze Σ_1. Prvky z $\mathcal{C}$ se nazývají ***kódová slova***.

Dekódovací pravidla

FI Buď dán kanál bez paměti se vstupní abecedou $\Sigma_1 = \{a_1, \ldots, a_m\}$ a výstupní abecedou $\Sigma_2 = \{b_1, \ldots, b_k\}$. ***Kód délky*** n je libovolný systém $\mathcal{C}$ různých posloupností délky n symbolů ze Σ_1. Prvky z $\mathcal{C}$ se nazývají ***kódová slova***.

Je-li dán kód $\mathcal{C}$ délky n s kódovými slovy $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$, ***dekódovací pravidlo*** je libovolný rozklad množiny možných obdržených posloupností do disjunktních množin $R_1, R_2, \ldots, R_N$ se zřejmou interpretací toho, že je-li obdržená posloupnost $\mathbf{y}$ prvkem množiny R_j, je $\mathbf{y}$ dekódováno jako kódové slovo $\mathbf{c}_j$.

Dekódovací pravidla

Formálně vzato, předpokládáme-li, že kód $\mathcal{C}$ neobsahuje např. symbol "?"jakožto kódové slovo, ***rozhodovací (dekódovací) pravidlo*** pro kód $\mathcal{C}$ je funkce $f : \Sigma_2^n \to \mathcal{C} \cup \{?\}$. Aplikaci dekódovacího pravidla nazýváme ***dekódování***.

Dekódovací pravidla

Formálně vzato, předpokládáme-li, že kód $\mathcal{C}$ neobsahuje např. symbol "?"jakožto kódové slovo, ***rozhodovací (dekódovací) pravidlo*** pro kód $\mathcal{C}$ je funkce $f : \Sigma_2^n \to \mathcal{C} \cup \{?\}$. Aplikaci dekódovacího pravidla nazýváme ***dekódování***.

Je-li $\mathbf{y}$ (obdržené) slovo v Σ_2^n, pak rozhodovací pravidlo *dekóduje* $\mathbf{y}$ jakožto kódové slovo $f(\mathbf{y})$ nebo v opačném případě nahlásí *dekódovací chybu*, jestliže $f(\mathbf{y}) = ?$.

Dekódovací pravidla

Formálně vzato, předpokládáme-li, že kód $\mathcal{C}$ neobsahuje např. symbol "?"jakožto kódové slovo, ***rozhodovací (dekódovací) pravidlo*** pro kód $\mathcal{C}$ je funkce $f : \Sigma_2^n \to \mathcal{C} \cup \{?\}$. Aplikaci dekódovacího pravidla nazýváme ***dekódování***.

Je-li $\mathbf{y}$ (obdržené) slovo v Σ_2^n, pak rozhodovací pravidlo *dekóduje* $\mathbf{y}$ jakožto kódové slovo $f(\mathbf{y})$ nebo v opačném případě nahlásí *dekódovací chybu*, jestliže $f(\mathbf{y}) = ?$.

Výběr dekódovacího pravidla je podstatný k úspěchu každého komunikačního systému. Jako extrémní příklad je snadné zkonstruovat dekódovací pravidlo, které zcela zničí bezchybnost kanálu bez šumu.

Příklad - dekódovací pravidla

Příklad 4.1

Předpokládejme, že máme zdroj s právě dvěma zdrojovými slovy s_1 a s_2, který můžeme zakódovat pro přenos binárním symetrickým kanálem jako

$$s_1 \mapsto 000 = \mathbf{c}_1, \qquad s_2 \mapsto 111 = \mathbf{c}_2.$$

Příklad - dekódovací pravidla

Příklad 4.1

Předpokládejme, že máme zdroj s právě dvěma zdrojovými slovy s_1 a s_2, který můžeme zakódovat pro přenos binárním symetrickým kanálem jako

$$s_1 \mapsto 000 = \mathbf{c}_1, \qquad s_2 \mapsto 111 = \mathbf{c}_2.$$

Máme pak osm možných obdržených zpráv. Možné dekódovací pravidlo by mohlo být dekódovat zprávu jako s_1, pokud obsahuje více nul než jedniček.

Příklad - dekódovací pravidla

Příklad 4.1

Předpokládejme, že máme zdroj s právě dvěma zdrojovými slovy s_1 a s_2, který můžeme zakódovat pro přenos binárním symetrickým kanálem jako

$$s_1 \mapsto 000 = \mathbf{c}_1, \qquad s_2 \mapsto 111 = \mathbf{c}_2.$$

Máme pak osm možných obdržených zpráv. Možné dekódovací pravidlo by mohlo být dekódovat zprávu jako s_1, pokud obsahuje více nul než jedniček.

Méně citlivé pravidlo by mohlo být dekódovat zprávu jako s_1, pouze když obdržená zpráva byla 000. A priori, každé z těchto pravidel má stejnou váhu, i s pravidlem: dekódujte každé obdržené slovo jako s_1!

Dekódovací pravidla

Naší snahou bude najít dekódovací pravidlo, které maximalizuje pravděpodobnost správného dekódování tj. pravděpodobnost, že $\mathbf{x} = f(\mathbf{y})$ je opravdu to kódové slovo $\mathbf{c}$, které bylo odesláno.

Dekódovací pravidla

Naší snahou bude najít dekódovací pravidlo, které maximalizuje pravděpodobnost správného dekódování tj. pravděpodobnost, že $\mathbf{x} = f(\mathbf{y})$ je opravdu to kódové slovo $\mathbf{c}$, které bylo odesláno.

Poznamenejme, že příjemce nemá žádnou možnost zjistit, zdali dekódovací proces opravdu dekódoval správně.

Dekódovací pravidla

Naší snahou bude najít dekódovací pravidlo, které maximalizuje pravděpodobnost správného dekódování tj. pravděpodobnost, že $\mathbf{x} = f(\mathbf{y})$ je opravdu to kódové slovo $\mathbf{c}$, které bylo odesláno.

Poznamenejme, že příjemce nemá žádnou možnost zjistit, zdali dekódovací proces opravdu dekódoval správně.

Pravděpodobnost správného dekódování lze vyjádřit mnoha způsoby. Použijeme-li například formuli úplné pravděpodobnosti, obdržíme následující dva vztahy:

$$P(\text{správné dekódování}) = \sum_{\mathbf{c}\in\mathcal{C}} P(\text{správné dekódování}|\mathbf{c}\text{ odesláno})P(\mathbf{c}\text{ odesláno}), \quad (4.1)$$

vztahujeme-li podmínku na množinu kódových slov resp.

Dekódovací pravidla

$$P(\text{správné dekódování}) = \sum_{\mathbf{y} \in \Sigma_2^n} P(\text{správné dekódování} | \mathbf{y} \text{ obdrženo}) P(\mathbf{y} \text{ obdrženo}), \tag{4.2}$$

vztahujeme-li podmínku na množinu slov ze Σ_2^n.

Dekódovací pravidla

$$P(\text{správné dekódování}) = \sum_{\mathbf{y} \in \Sigma_2^n} P(\text{správné dekódování} | \mathbf{y} \text{ obdrženo}) P(\mathbf{y} \text{ obdrženo}), \tag{4.2}$$

vztahujeme-li podmínku na množinu slov ze Σ_2^n.

Poznamenejme, že vztah (4.1) explicitně obsahuje pravděpodobnosti $P(\mathbf{c}$ odesláno), že různá kódová slova byla poslána pomocí kanálu.

Dekódovací pravidla

$$P(\text{správné dekódování}) = \sum_{\mathbf{y}\in\Sigma_2^n} P(\text{správné dekódování}|\mathbf{y}\text{ obdrženo})P(\mathbf{y}\text{ obdrženo}), \tag{4.2}$$

vztahujeme-li podmínku na množinu slov ze Σ_2^n.

Poznamenejme, že vztah (4.1) explicitně obsahuje pravděpodobnosti $P(\mathbf{c}$ odesláno), že různá kódová slova byla poslána pomocí kanálu.

Tyto pravděpodobnosti nejsou nic jiného než pravděpodobnosti zdroje $\mathcal{C}$. Mluvíme pak o ***vstupním rozdělení kanálu***.

Dekódovací pravidla

$$P(\text{správné dekódování}) = \sum_{\mathbf{y}\in\Sigma_2^n} P(\text{správné dekódování}|\mathbf{y}\text{ obdrženo})P(\mathbf{y}\text{ obdrženo}), \tag{4.2}$$

vztahujeme-li podmínku na množinu slov ze Σ_2^n.

Poznamenejme, že vztah (4.1) explicitně obsahuje pravděpodobnosti $P(\mathbf{c}$ odesláno), že různá kódová slova byla poslána pomocí kanálu.

Tyto pravděpodobnosti nejsou nic jiného než pravděpodobnosti zdroje $\mathcal{C}$. Mluvíme pak o ***vstupním rozdělení kanálu***.

Přitom (4.2) rovněž obsahuje vstupní rozdělení, protože pravděpodobnost, že dané slovo $\mathbf{y}$ je obdrženo, obvykle závisí na tom, které kódové slovo bylo odesláno.

Dekódovací pravidla

Nechť f je dekódovací pravidlo pro kód $\mathcal{C}$. Je-li odesláno kódové slovo $\mathbf{c}$, pak správné dekódování nastane právě tehdy, když $f(\mathbf{y}) = \mathbf{c}$ pro obdržené slovo $\mathbf{y}$. Platí tedy

$$P(\text{správné dekódování}|\mathbf{c}\text{ odesláno}) = \sum_{\mathbf{y}\in\Sigma_2^n, f(\mathbf{y})=\mathbf{c}} P(\mathbf{y}\text{ obdrženo}|\mathbf{c}\text{ odesláno}). \quad (4.3)$$

Dekódovací pravidla

Nechť f je dekódovací pravidlo pro kód $\mathcal{C}$. Je-li odesláno kódové slovo $\mathbf{c}$, pak správné dekódování nastane právě tehdy, když $f(\mathbf{y}) = \mathbf{c}$ pro obdržené slovo $\mathbf{y}$. Platí tedy

$$P(\text{správné dekódování}|\mathbf{c}\text{ odesláno}) = \sum_{\mathbf{y}\in\Sigma_2^n, f(\mathbf{y})=\mathbf{c}} P(\mathbf{y}\text{ obdrženo}|\mathbf{c}\text{ odesláno}). \quad (4.3)$$

Provedeme-li substituci do (4.1), obdržíme

$$P(\text{správné dekódování}) = \sum_{\mathbf{c}\in\mathcal{C}, \mathbf{y}\in\Sigma_2^n, f(\mathbf{y})=\mathbf{c}} P(\mathbf{y}\text{ obdrženo}|\mathbf{c}\text{ odesláno})P(\mathbf{c}\text{ odesláno}). \quad (4.4)$$

Dekódovací pravidla

Tato dvojnásobná suma není však vždy zcela příhodná. Přitom vztah (4.2) nám podává vhodnější návod, jak obdržet dobré dekódovací pravidlo.

Dekódovací pravidla

Tato dvojnásobná suma není však vždy zcela příhodná. Přitom vztah (4.2) nám podává vhodnější návod, jak obdržet dobré dekódovací pravidlo.

Podle dekódovacího pravidla f je obdržené slovo $\mathbf{y}$ dekódováno správně, jestliže odeslané slovo bylo $f(\mathbf{y})$.

Dekódovací pravidla

Tato dvojnásobná suma není však vždy zcela příhodná. Přitom vztah (4.2) nám podává vhodnější návod, jak obdržet dobré dekódovací pravidlo.

Podle dekódovacího pravidla f je obdržené slovo $\mathbf{y}$ dekódováno správně, jestliže odeslané slovo bylo $f(\mathbf{y})$.

Platí tedy

$$P(\text{správné dekódování}|\ \mathbf{y}\ \text{obdrženo}) = P(f(\mathbf{y})\ \text{odesláno}|\ \mathbf{y}\ \text{obdrženo}) \quad (4.5)$$

a přitom se ve výše uvedeném výrazu nevyskytuje žádná suma.

Dekódovací pravidla

Dosaďme do vztahu (4.2). Pak máme

$$P(\text{správné dekódování}) = \sum_{\mathbf{y}\in\Sigma_2^n} P(f(\mathbf{y}) \text{ odesláno}|\mathbf{y} \text{ obdrženo})P(\mathbf{y} \text{ obdrženo}). \quad (4.6)$$

Dekódovací pravidla

Dosaďme do vztahu (4.2). Pak máme

$$P(\text{správné dekódování}) = \sum_{\mathbf{y}\in\Sigma_2^n} P(f(\mathbf{y}) \text{ odesláno}|\mathbf{y} \text{ obdrženo})P(\mathbf{y} \text{ obdrženo}). \quad (4.6)$$

Pravděpodobnost správného kódování lze maximalizovat tím, že budeme postupovat podle takového dekódovacího pravidla, které maximalizuje každou z podmíněných pravděpodobností

$$P(f(\mathbf{y}) \text{ odesláno}|\mathbf{y} \text{ obdrženo}).$$

Pravidlo ideálního pozorovatele

Jinak řečeno, za předpokladu, že jsme obdrželi $\mathbf{y}$, rozhodneme se tak, že kódové slovo, které bylo posláno, je to nejpravděpodobnější, které mohlo být odesláno. To jde konkrétně zajistit tak, že se procházíme zpětnými kanálovými pravděpodobnostmi

$$P(\mathbf{c}_1 \text{ odesláno}|\mathbf{y} \text{ obdrženo}), \ldots, P(\mathbf{c}_N \text{ odesláno}|\mathbf{y} \text{ obdrženo})$$

a vybereme kódové slovo $\mathbf{c}_i$ s největší pravděpodobností.

Pravidlo ideálního pozorovatele

Jinak řečeno, za předpokladu, že jsme obdrželi $\mathbf{y}$, rozhodneme se tak, že kódové slovo, které bylo posláno, je to nejpravděpodobnější, které mohlo být odesláno. To jde konkrétně zajistit tak, že se procházíme zpětnými kanálovými pravděpodobnostmi

$$P(\mathbf{c}_1 \text{ odesláno}|\mathbf{y} \text{ obdrženo}), \ldots, P(\mathbf{c}_N \text{ odesláno}|\mathbf{y} \text{ obdrženo})$$

a vybereme kódové slovo $\mathbf{c}_i$ s největší pravděpodobností.

Toto pravidlo se nazývá pravidlo ***ideálního pozorovatele*** neboli ***pravidlo minimální chyby***.

Pravidlo ideálního pozorovatele

Nicméně, přepis těchto podmíněných pravděpodobností nám ukazuje, že tyto podmíněné pravděpodobnosti nelze použít bez znalosti pravděpodobností výskytu kódových slov $\mathbf{c}_j$.
Máme totiž podle Bayesovy věty:

$$P(\mathbf{c} \text{ odesláno}|\mathbf{y} \text{ obdrženo}) = \frac{P(\mathbf{y} \text{ obdrženo}|\mathbf{c} \text{ odesláno})P(\mathbf{c} \text{ odesláno})}{\sum_{k=1}^{N} P(\mathbf{y} \text{ obdrženo}|\mathbf{c}_k \text{ odesláno})P(\mathbf{c}_k \text{ odesláno})}. \tag{4.7}$$

Pravidlo ideálního pozorovatele

Nicméně, přepis těchto podmíněných pravděpodobností nám ukazuje, že tyto podmíněné pravděpodobnosti nelze použít bez znalosti pravděpodobností výskytu kódových slov $\mathbf{c}_j$.
Máme totiž podle Bayesovy věty:

$$P(\mathbf{c} \text{ odesláno}|\mathbf{y} \text{ obdrženo}) = \frac{P(\mathbf{y} \text{ obdrženo}|\mathbf{c} \text{ odesláno})P(\mathbf{c} \text{ odesláno})}{\sum_{k=1}^{N} P(\mathbf{y} \text{ obdrženo}|\mathbf{c}_k \text{ odesláno})P(\mathbf{c}_k \text{ odesláno})}. \tag{4.7}$$

V praxi je to vážná nevýhoda. Totiž, abychom určili dekódovací funkci, musíme znát, s jakou pravděpodobností jsou kódová slova posílána pomocí kanálu, tj. musíme znát jistou informaci o zprávě, což není zrovna vždy možné.

Pravidlo maximální pravděpodobnosti

Toto, společně se skutečností, že není snadné toto pravidlo aplikovat v případě, kdy máme velký počet kódových slov, opravňuje užití následujícího pravidla nazývaného ***pravidlem maximální pravděpodobnosti*** (***maximum-likelihood (ML)***).

Pravidlo maximální pravděpodobnosti

Toto, společně se skutečností, že není snadné toto pravidlo aplikovat v případě, kdy máme velký počet kódových slov, opravňuje užití následujícího pravidla nazývaného ***pravidlem maximální pravděpodobnosti*** (***maximum-likelihood (ML)***).

Toto pravidlo dekóduje každý obdržený vektor $\mathbf{y}$ do kódového slova $\mathbf{c}_j$ tak, že maximalizuje

$$P(\mathbf{y} \text{ obdrženo} | \mathbf{c}_j \text{ odesláno}). \tag{4.8}$$

Pro ty, kteří jsou obeznámeni s odhady maximální pravděpodobnosti ve statistice, je analogie zřejmá.

Splynutí obou pravidel

Za předpokladu nedostatku informace o pravděpodobnostech různých kódových slov máme následující:

Mají-li kódová slova stejnou pravděpodobnost, pak pravidlo maximální pravděpodobnosti splývá s pravidlem ideálního pozorovatele. (4.9)

Splynutí obou pravidel

Za předpokladu nedostatku informace o pravděpodobnostech různých kódových slov máme následující:

Mají-li kódová slova stejnou pravděpodobnost, pak pravidlo maximální pravděpodobnosti splývá s pravidlem ideálního pozorovatele. (4.9)

Důkaz je snadný. Totiž platí $P(\mathbf{c} \text{ odesláno}) = \frac{1}{N}$. Tedy platí dle (4.7)

$$P(\mathbf{c} \text{ odesláno}|\mathbf{y} \text{ obdrženo}) = \frac{P(\mathbf{y} \text{ obdrženo}|\mathbf{c} \text{ odesláno})}{\sum_{k=1}^{N} P(\mathbf{y} \text{ obdrženo}|\mathbf{c}_k \text{ odesláno})}. \quad (4.10)$$

Splynutí obou pravidel

Za předpokladu nedostatku informace o pravděpodobnostech různých kódových slov máme následující:

Mají-li kódová slova stejnou pravděpodobnost, pak pravidlo maximální pravděpodobnosti splývá s pravidlem ideálního pozorovatele. (4.9)

Důkaz je snadný. Totiž platí $P(\mathbf{c} \text{ odesláno}) = \frac{1}{N}$. Tedy platí dle (4.7)

$$P(\mathbf{c} \text{ odesláno}|\mathbf{y} \text{ obdrženo}) = \frac{P(\mathbf{y} \text{ obdrženo}|\mathbf{c} \text{ odesláno})}{\sum_{k=1}^{N} P(\mathbf{y} \text{ obdrženo}|\mathbf{c}_k \text{ odesláno})}. \quad (4.10)$$

Odtud pak máme, že maximum na pravé straně obdržíme právě tehdy, když budeme mít maximum na levé straně.

Hammingova vzdálenost

V hlavní části této přednášky budeme pracovat s binárním symetrickým kanálem. Pro tento kanál má pravidlo maximální pravděpodobnosti obzvlášť snadnou implementaci.

Hammingova vzdálenost

V hlavní části této přednášky budeme pracovat s binárním symetrickým kanálem. Pro tento kanál má pravidlo maximální pravděpodobnosti obzvlášť snadnou implementaci.

Nechť V_n označuje množinu všech posloupností délky n složených z nul a jedniček a, pokud to bude nutné, považujme V_n za vektorový n-dimenzionální prostor nad tělesem celých čísel modulo 2.

Hammingova vzdálenost

V hlavní části této přednášky budeme pracovat s binárním symetrickým kanálem. Pro tento kanál má pravidlo maximální pravděpodobnosti obzvlášť snadnou implementaci.

Nechť V_n označuje množinu všech posloupností délky n složených z nul a jedniček a, pokud to bude nutné, považujme V_n za vektorový n-dimenzionální prostor nad tělesem celých čísel modulo 2.

Jsou-li $\mathbf{x}$ a $\mathbf{y}$ vektory z V_n, definujme ***Hammingovu vzdálenost*** $d(\mathbf{x}, \mathbf{y})$ mezi $\mathbf{x}$ a $\mathbf{y}$ jako počet míst, ve kterých se $\mathbf{x}$ a $\mathbf{y}$ liší.

Pravidlo minimální vzdálenosti

Pro binární symetrický kanál je přirozeným dekódovacím pravidlem pravidlo ***minimální vzdálenosti***, totiž: dekódujme každý obdržený vektor $\mathbf{y}$ do kódového slova $\mathbf{c}_j$, které má minimální Hammingovu vzdálenost od $\mathbf{y}$: pokud je vícero takových slov, vybereme $\mathbf{c}_j$ libovolně.

$$P(\mathbf{y} \text{ obdrženo} | \mathbf{c}_j \text{ odesláno}). \tag{4.11}$$

Pravidlo minimální vzdálenosti

Pro binární symetrický kanál je přirozeným dekódovacím pravidlem pravidlo ***minimální vzdálenosti***, totiž: dekódujme každý obdržený vektor $\mathbf{y}$ do kódového slova $\mathbf{c}_j$, které má minimální Hammingovu vzdálenost od $\mathbf{y}$: pokud je vícero takových slov, vybereme $\mathbf{c}_j$ libovolně.

$$P(\mathbf{y} \text{ obdrženo} | \mathbf{c}_j \text{ odesláno}). \tag{4.11}$$

Následující snadný výsledek nám tvrdí:

Věta 4.2

Pro binární symetrický kanál s pravděpodobností chyby $p \leq \frac{1}{2}$ je dekódovací pravidlo minimální vzdálenosti ekvivalentní k pravidlu maximální pravděpodobnosti.

Důkaz Věty 4.2

Pro všechny vektory $\mathbf{x}$ a $\mathbf{y}$ z V_n s vlastností $d(\mathbf{x}, \mathbf{y}) = d$ platí

$$P(\mathbf{y} \text{ bylo obdrženo} | \mathbf{x} \text{ bylo odesláno}) = p^d q^{n-d}.$$

Důkaz Věty 4.2

Pro všechny vektory $\mathbf{x}$ a $\mathbf{y}$ z V_n s vlastností $d(\mathbf{x}, \mathbf{y}) = d$ platí

$$P(\mathbf{y} \text{ bylo obdrženo} | \mathbf{x} \text{ bylo odesláno}) = p^d q^{n-d}.$$

Pokud $p < \frac{1}{2}$, tento výraz nabývá maxima, je-li d minimální. To ale zřejmě stačí k tomu, že pevné slovo $\mathbf{y}$ dekódujeme jako to kódové slovo, které má nejmenší vzdálenost od slova $\mathbf{y}$.

Důkaz Věty 4.2

Pro všechny vektory $\mathbf{x}$ a $\mathbf{y}$ z V_n s vlastností $d(\mathbf{x}, \mathbf{y}) = d$ platí

$$P(\mathbf{y} \text{ bylo obdrženo} | \mathbf{x} \text{ bylo odesláno}) = p^d q^{n-d}.$$

Pokud $p < \frac{1}{2}$, tento výraz nabývá maxima, je-li d minimální. To ale zřejmě stačí k tomu, že pevné slovo $\mathbf{y}$ dekódujeme jako to kódové slovo, které má nejmenší vzdálenost od slova $\mathbf{y}$.

Obráceně, vezmeme-li jako rozkódování pevného slova $\mathbf{y}$ kódové slovo minimální vzdálenosti, je výše uvedená pravděpodobnost maximální.

Cvičení

Cvičení 4.3

1. *Nechť kód sestává ze čtyř kódových slov* $\mathbf{c}_1 = 1000$, $\mathbf{c}_2 = 0110$, $\mathbf{c}_3 = 0001$ *a* $\mathbf{c}_4 = 1111$. *Pravděpodobnosti výskytu těchto kódových slov jsou dány jako*

$$P(\mathbf{c}_1) = P(\mathbf{c}_2) = \frac{1}{3}, \quad P(\mathbf{c}_3) = P(\mathbf{c}_4) = \frac{1}{6}$$

Používáte-li pro přenos binární symetrický kanál s pravděpodobnosti chyby $\frac{1}{10}$ *a obdržíte na výstup vektor* 1001, *jak by jste se rozhodoval při*

   1. *použití pravidla ideálního pozorovatele,*
   2. *použitím pravidla maximální pravděpodobnosti?*

Obsah

Kapacita kanálu - Definice

FI Jak už napovídá jméno, kapacita komunikačního kanálu je míra jeho schopnosti přenášet informaci. Formální definice je motivována níže uvedeným:
Předpokládejme, že máme diskrétní kanál bez paměti se vstupní abecedou $\Sigma_1 = \{a_1, \ldots, a_m\}$, výstupní abecedou $\Sigma_2 = \{b_1, \ldots, b_n\}$ a maticí P kanálu

$$P = [p_{ij}] = P(\mathbf{b}_j \text{ obdrženo} | \mathbf{a}_i \text{ odesláno}).$$

Kapacita kanálu - Definice

FI Jak už napovídá jméno, kapacita komunikačního kanálu je míra jeho schopnosti přenášet informaci. Formální definice je motivována níže uvedeným:
Předpokládejme, že máme diskrétní kanál bez paměti se vstupní abecedou $\Sigma_1 = \{a_1, \ldots, a_m\}$, výstupní abecedou $\Sigma_2 = \{b_1, \ldots, b_n\}$ a maticí P kanálu

$$P = [p_{ij}] = P(\mathbf{b}_j \text{ obdrženo} | \mathbf{a}_i \text{ odesláno}).$$

Přidáme-li k tomuto kanálu zdroj $\mathcal{S}$ bez paměti, který vysílá symboly $a_1, \ldots, a_m$ s pravděpodobnostmi $p_1, \ldots, p_m$, pak výstup kanálu můžeme považovat za zdroj $\mathcal{J}$ bez paměti.

Kapacita kanálu - Definice

Ten vysílá symboly $b_1, \ldots, b_n$ s pravděpodobnostmi $q_1, \ldots, q_n$,

kde $$\begin{aligned} q_j &= \sum_{i=1}^{m} P(\mathbf{b}_j \text{ obdrženo} | \mathbf{a}_i \text{ odesláno}) P(\mathbf{a}_i \text{ odesláno}) \\ &= \sum_{i=1}^{m} p_i p_{ij}. \end{aligned}$$

Kapacita kanálu - Definice

Ten vysílá symboly $b_1, \ldots, b_n$ s pravděpodobnostmi $q_1, \ldots, q_n$, kde $q_j = \sum_{i=1}^{m} P(\mathbf{b}_j \text{ obdrženo} | \mathbf{a}_i \text{ odesláno}) P(\mathbf{a}_i \text{ odesláno})$
$= \sum_{i=1}^{m} p_i p_{ij}$.

Informace o $\mathcal{S}$ podaná pomocí $\mathcal{J}$, definovaná v kapitole 1, je rovna

$$I(\mathcal{S}|\mathcal{J}) = H(\mathcal{S}) - H(\mathcal{S}|\mathcal{J}) = H(\mathcal{S}) + H(\mathcal{J}) - H(\mathcal{S}, \mathcal{J})$$

a je to funkce, která závisí pouze na pravděpodobnostním rozdělení $p_1, \ldots, p_m$, a matici kanálu P. Je proto přirozené definovat *kapacitu* C kanálu jako

$$C = \sup I(\mathcal{S}|\mathcal{J}), \tag{5.1}$$

kde supremum je bráno přes všechny zdroje bez paměti $\mathcal{S}$, nebo, ještě přesněji, nad všemi možnými rozděleními pravděpodobností $(p_1, \ldots, p_n)$.

Kapacita kanálu - Definice

Nejdříve si připomeňme, že C je dobře definováno v tom smyslu, že pouze hledáme supremum funkce $f(\mathbf{p})$, kde f je spojitá funkce na uzavřené a ohraničené podmnožině množiny $\mathbf{R}^m$ a dle základní věty analýzy má f maximum v nějakém bodě. Můžeme tedy 5.1 přepsat jako

$$C = \max I(\mathcal{S}|\mathcal{T}), \tag{5.2}$$

Dále si uvědomme, že C je kvantitativní veličina určená pouze maticí kanálu P. Můžeme ji zhruba považovat za konduktanci (vodivost) odporu v teorii elektrických obvodů. Její jednotky jsou pak jednotky informace nebo entropie, totiž "bity za sekundu" nebo "bity na symbol" v závislosti na kontextu.

Kapacita binárního symetrického kanálu

Ukažme příklad, jak lze najít kapacitu kanálu.

Věta 5.1

Kapacita binárního symetrického kanálu s pravděpodobností chyby přenosu p je určena vztahem

$$C(p) = 1 + p \log p + q \log q, \tag{5.3}$$

kde $q = 1 - p$.

Kapacita - Důkaz Věty 5.1

Důkaz.

K usnadnění označení předpokládejme, že zdroj $\mathcal{S}$ emituje 0 s pravděpodobností α a 1 s pravděpodobností $\beta = 1 - \alpha$. Pak výstup $\mathcal{J}$ má rozdělení

$$0 \text{ s pravděpodobností } \alpha q + \beta p, \quad 1 \text{ s pravděpodobností } \beta q + \alpha p.$$

Je tedy $H(\mathcal{S}, \mathcal{J})$ právě entropie rozdělení $(\alpha q, \alpha p, \beta q, \beta q)$. Po jednoduché úpravě

$$\begin{aligned} I(\mathcal{S}|\mathcal{J}) &= p \log p + q \log q - (\alpha q + \beta p) \log(\alpha q + \beta p) \\ &\quad -(\alpha p + \beta q) \log(\alpha p + \beta q) \end{aligned}$$

Derivujme dle α. Pak obdržíme, že $I(\mathcal{S}|\mathcal{J})$ má maximum v případě, že $\alpha = \frac{1}{2}$ a obdržíme pak 5.3. ∎

Vlastnosti kapacity

Poznamenejme, že kapacita má očekávané vlastnosti – $C(p)$ je monotonní funkce p, $0 \leq p \leq \frac{1}{2}$, a

$$C(0) = 1, \qquad C(\frac{1}{2}) = 0,$$

což odpovídá intuici, že, pokud $p = \frac{1}{2}$, kanál se stane dokonalým rušičem, ale že, pokud $p = 0$, máme dokonalý přenos.

Vlastnosti kapacity

Poznamenejme, že kapacita má očekávané vlastnosti – $C(p)$ je monotonní funkce p, $0 \leq p \leq \frac{1}{2}$, a

$$C(0) = 1, \qquad C(\frac{1}{2}) = 0,$$

což odpovídá intuici, že, pokud $p = \frac{1}{2}$, kanál se stane dokonalým rušičem, ale že, pokud $p = 0$, máme dokonalý přenos.

Zjištění kapacity obecných kanálů je netriviální záležitost. V případě, že kanál nemá nějakou speciální vlastnost nebo není odvozen z kanálu, jehož kapacita je známa, jediný způsob, jak můžeme vypočítat kapacitu, je vyřešení problému optimalizace s omezeními, a to zejména metodou Lagrangeových multiplikátorů.

Kapacita *r*-tého rozšíření

Příkladem první z těchto technik je následující výsledek.

Věta 5.2

Má-li kanál $\mathcal{S}$ bez paměti kapacitu C, má jeho r-té rozšíření $\mathcal{S}^{(r)}$ kapacitu rC.

Důkaz.

Označme jako $C^{(r)}$ kapacitu $r-$tého rozšíření tak, že

$$C^{(r)} = \sup_{\mathbf{X}} H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}), \tag{5.4}$$

kde $\mathbf{X} = (X_1, \ldots, X_r)$ a $\mathbf{Y} = (Y_1, \ldots, Y_r)$ jsou vstupní a výstupní dvojice. Máme ale

$$H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}). \tag{5.5}$$

Důkaz Věty 5.2

Pokračování důkazu.

Zároveň

$$H(\mathbf{Y}|\mathbf{X}) = \sum_{\mathbf{x}} p(\mathbf{x}) H(\mathbf{Y}|\mathbf{X} = \mathbf{x}).$$

Protože se jedná o kanál bez paměti, máme

$$H(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = \sum_i H(Y_i|\mathbf{X} = \mathbf{x}) = \sum_i H(Y_i|X_i = x_i).$$

Zejména

$$\begin{aligned} H(\mathbf{Y}|\mathbf{X}) &= \textstyle\sum_{\mathbf{x}} p(\mathbf{x}) H(Y_i|X_i = x_i) \\ &= \textstyle\sum_i \sum_u H(Y_i|X_i = u) \cdot P(X_i = u). \end{aligned}$$

∎

Důkaz Věty 5.2

Pokračování důkazu.

Tedy

$$H(\mathbf{Y}|\mathbf{X}) = \sum_{i}^{r} H(Y_i|X_i). \tag{5.6}$$

Obecně platí

$$H(\mathbf{Y}) \leq H(Y_1) + \cdots + H(Y_r),$$

a tedy celkem $C^{(r)} \leq rC$. Připomeňme, že rovnost nastává právě tehdy, když $Y_1, \ldots, Y_r$ jsou nezávislá. Toho lze dosáhnout tím, že zvolíme $X_1, \ldots, X_r$ jako nezávislé a vybráním rozdělení, při kterém bylo dosaženo kapacity C kanálu. ∎

Cvičení

Cvičení 5.3

1. *Vypočtěte kapacitu binárního vypouštěcího kanálu s pravděpodobností chyby ε.*
2. *Uvažujeme-li kanál bez paměti s maticí*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},$$

ukažte, že kapacity lze dosáhnout více než jedním rozdělením na vstupu. Ukažte, že rozšířením 2. řádu můžeme dosáhnout kapacity pomocí rozdělení na vstupu, které není součinem rozdělení na vstupu původního kanálu.

(Feinstein, 1958)

Obsah

Věta o kódování se šumem

FI Již dříve jsme viděli, že můžeme dosáhnou libovolně velké spolehlivosti pouze dostatečně častým opakováním každého zdrojového symbolu. Zřejmě je tato metoda velmi časově náročná a hlavním účelem tohoto odstavce je dokázat překrásné tvrzení C. Shannona (1948), které tvrdí, že za předpokladu, že rychlost (míra) přenosu je pod kapacitou kanálu, lze dosáhnout libovolně velké spolehlivosti.

Věta o kódování se šumem

FI Již dříve jsme viděli, že můžeme dosáhnou libovolně velké spolehlivosti pouze dostatečně častým opakováním každého zdrojového symbolu. Zřejmě je tato metoda velmi časově náročná a hlavním účelem tohoto odstavce je dokázat překrásné tvrzení C. Shannona (1948), které tvrdí, že za předpokladu, že rychlost (míra) přenosu je pod kapacitou kanálu, lze dosáhnout libovolně velké spolehlivosti.

Budeme se koncentrovat na binární symetrický kanál. Tyto myšlenky lze rozšířit na podstatně komplikovanější kanály, ale důležitější je plně porozumět nosným principům, než se obklopit matematickými detaily.

Věta o kódování se šumem

Buď dán kód $\mathcal{C}$ a dekódovací schéma pro $\mathcal{C}$.

Pravděpodobnost chyby $\mathrm{e}(\mathcal{C})$ je obvykle definovaná jako průměrná pravděpodobnost chyby za předpokladu, že všechna kódová slova byla vyslána se stejnou pravděpodobností.

Jinak řečeno, máme-li M kódových slov $\mathbf{c}_1, \ldots, \mathbf{c}_M$ z $\mathcal{C}$, pak platí

$$\mathrm{e}(\mathcal{C}) = \frac{1}{M} \sum_{i=1}^{M} P(\text{nastala chyba} | \mathbf{c}_i \text{ bylo přeneseno}).$$

V případě binárních kódů můžeme předpokládat, pokud nebude jinak uvedeno, že používáme dekódovací pravidlo maximální pravděpodobnosti (=pravidlo minimální vzdálenosti), a tudíž se často budeme odvolávat na pravděpodobnost chyby kódování bez specifického připomenutí dekódovacího pravidla.

Věta o kódování se šumem

Budeme se snažit najít kódy s malou průměrnou pravděpodobností chyby. Avšak, podstatně silnějším požadavkem je, aby ***maximální pravděpodobnost chyby*** je malá. Jak lze očekávat, ta je definována jako

$$\hat{e}(\mathcal{C}) = \max_i P(\text{nastala chyba} | \mathbf{c}_i \text{ bylo přeneseno}),$$

a evidentně

$$\hat{e} \geq e.$$

Předpokládejme proto, že máme binární symetrický kanál s pravděpodobností chyby p a tudíž kapacitou C určenou

$$C = C(p) = 1 + p \log p + (1 - p) \log (1 - p).$$

Věta o kódování se šumem

Dokažme následující verzi Shannonovy věty.

Theorem 6.1

Shannonova věta o kódování se šumem *Buď dán binární symetrický kanál kapacity C a libovolné R, $0 < R < C$. Pak pro každou posloupnost $(M_n : 1 \leq n < \infty)$ přirozených čísel splňujících*

$$1 \leq M_n \leq 2^{Rn} \quad (1 \leq n < \infty),$$

a všechna kladná $\varepsilon > 0$, existuje posloupnost kódů $(\mathcal{C}_n : 1 \leq n < \infty)$ a přirozené číslo $N_0(\varepsilon)$ tak, že $\mathcal{C}_n$ má M_n kódových slov délky n a maximální pravděpodobnost chyby

$$\hat{e}(\mathcal{C}_n) \leq \varepsilon$$

pro všechna $n \geq N_0(\varepsilon)$.

Věta o kódování se šumem

Jakým způsobem funguje tato věta. Předpokládejme, že pravděpodobnost chyby takovéhoto kanálu je taková, že kapacita kanálu $C(p) = 0.8$. Pak, je-li naše zpráva řetězec nul a jedniček, víme, že pro dostatečně velké n, položíme-li $R = 0.75$, existuje množina $2^{0.75n}$ kódových slov délky n, která mají pravděpodobnost chyby menší než libovolně předem předepsaná hranice.

Věta o kódování se šumem

Jakým způsobem funguje tato věta. Předpokládejme, že pravděpodobnost chyby takovéhoto kanálu je taková, že kapacita kanálu $C(p) = 0.8$. Pak, je-li naše zpráva řetězec nul a jedniček, víme, že pro dostatečně velké n, položíme-li $R = 0.75$, existuje množina $2^{0.75n}$ kódových slov délky n, která mají pravděpodobnost chyby menší než libovolně předem předepsaná hranice. Tudíž, abychom zakódovali zprávu ze zdroje, postup je následující:

(a) Rozdělte zprávu do bloků délky m, přičemž m je takové, že $3\lceil \frac{n}{4} \rceil = m \geq N_0(\varepsilon)$.

(b) Zakódujte každý z těchto m-bloků do kódu $\mathcal{C}_n$ tak, že použijete kódové slovo délky $\frac{4m}{3}$ pro každý m-blok.

(b) Přeneste nově zakódovanou posloupnost kanálem.

Věta o kódování se šumem

Čeho jsme dosáhli? Podstatné redukce pravděpodobnosti chyby. Na čí náklady? Komplexnosti zakódování a menší míry přenosu: zároveň však bohužel doposud neznámé zakódování. Síla Shannonovy věty spočívá v tom, že existují takovéto kódy.

Věta o kódování se šumem

Čeho jsme dosáhli? Podstatné redukce pravděpodobnosti chyby. Na čí náklady? Komplexnosti zakódování a menší míry přenosu: zároveň však bohužel doposud neznámé zakódování. Síla Shannonovy věty spočívá v tom, že existují takovéto kódy.

MA Důkaz Shannonovy věty, který chceme provést níže, závisí na dvou nerovnostech – z nich první je velmi dobře známa – její důkaz lze najít v každém elementárním textu z teorie pravděpodobnosti.

Čebyševova nerovnost

Je-li X libovolná náhodná proměnná tak, že má konečnou variaci (odchylku) $\mathrm{var}(X) = D(X)$, pak pro každé $a > 0$ máme

$$P(|X - E(X)| \geq a) \leq D(X)/a^2. \qquad (6.1)$$

Věta o kódování se šumem

Druhá nerovnost je méně známá a má rovněž pravděpodobnostní interpretaci; lze ji vyslovit následovně.

Omezená nerovnost

Pro všechna λ, $0 \leq \lambda \leq \frac{1}{2}$, platí

$$\sum_{k=0}^{\lfloor \lambda n \rfloor} \binom{n}{k} \leq 2^{n\mathrm{h}(\lambda)}, \tag{6.2}$$

kde $\mathrm{h}(\lambda) = -[\lambda \log \lambda + (1-\lambda) \log (1-\lambda)]$.

Důkaz Omezené nerovnosti

Důkaz.

Položme $m = \lfloor \lambda n \rfloor$.

Platí $\frac{\lambda}{1-\lambda} \leq 1$. Tedy pro $0 \leq k \leq m \leq \lambda n$ máme

$$\left(\frac{\lambda}{1-\lambda}\right)^{\lambda n} \leq \left(\frac{\lambda}{1-\lambda}\right)^{m} \leq \left(\frac{\lambda}{1-\lambda}\right)^{k} \leq 1.$$

Pak můžeme psát

$$\begin{aligned} 1 = [\lambda + (1-\lambda)]^n &\geq \sum_{k=0}^{m} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \\ &= (1-\lambda)^n \sum_{k=0}^{m} \binom{n}{k} \left(\frac{\lambda}{1-\lambda}\right)^k \\ &\geq \lambda^{\lambda n} (1-\lambda)^{n(1-\lambda)} \sum_{k=0}^{m} \binom{n}{k}. \end{aligned}$$

Důkaz Omezené nerovnosti

Pokračování.

Tudíž

$$\sum_{k=0}^{\lambda n} \binom{n}{k} \leq \lambda^{\lambda n}(1-\lambda)^{n(1-\lambda)} = 2^{n\mathrm{h}(\lambda)},$$

logaritmujeme-li při základu 2 a pak znovu umocníme. ∎

Důkaz věty o kódování se šumem

Důkaz Věty 6.1.

Nejprve popišme hrubý směr důkazu.

Důkaz věty o kódování se šumem

Důkaz Věty 6.1.

Nejprve popišme hrubý směr důkazu.

Zvolme si pevné přirozené číslo n, a pro daný okamžik, pracujme s binárními kódy ve V_n. Předpokládejme, že se pokoušíme najít kód s M kódovými slovy $\mathbf{c}_i \in V_n$.

Důkaz věty o kódování se šumem

Důkaz Věty 6.1.

Nejprve popišme hrubý směr důkazu.

Zvolme si pevné přirozené číslo n, a pro daný okamžik, pracujme s binárními kódy ve V_n. Předpokládejme, že se pokoušíme najít kód s M kódovými slovy $\mathbf{c}_i \in V_n$.

Vybereme ta kódová slova $\mathbf{c}_i$ trochu bláznivou metodou vybráním vektorů z V_n náhodně a nezávisle na i, $(1 \leq i \leq M)$. Tomuto kódování říkáme ***náhodné kódování***.

Důkaz věty o kódování se šumem

Důkaz Věty 6.1.

Nejprve popišme hrubý směr důkazu.

Zvolme si pevné přirozené číslo n, a pro daný okamžik, pracujme s binárními kódy ve V_n. Předpokládejme, že se pokoušíme najít kód s M kódovými slovy $\mathbf{c}_i \in V_n$.

Vybereme ta kódová slova $\mathbf{c}_i$ trochu bláznivou metodou vybráním vektorů z V_n náhodně a nezávisle na i, $(1 \leq i \leq M)$. Tomuto kódování říkáme ***náhodné kódování***.

Budeme kódovat následujícím způsobem: zvolme $r > 0$ a nechť $S_r(\mathbf{y})$ definuje *r-sféru* se středem $\mathbf{y}$, tj.

$$S_r(\mathbf{y}) = \{\mathbf{z} : \mathbf{z} \in V_n, d(\mathbf{y},\mathbf{z}) \leq r\}.$$

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Pak, je-li $\mathbf{y}$ obdržený vektor, můžeme dekódovat $\mathbf{y}$ jako kódové slovo $\mathbf{c}_j$, je-li $\mathbf{c}_j$ jediné kódové slovo v $S_r(\mathbf{y})$; jinak budeme dekódovat $\mathbf{y}$ jako libovolné jiné kódové slovo, např. $\mathbf{c}_1$.

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Pak, je-li $\mathbf{y}$ obdržený vektor, můžeme dekódovat $\mathbf{y}$ jako kódové slovo $\mathbf{c}_j$, je-li $\mathbf{c}_j$ jediné kódové slovo v $S_r(\mathbf{y})$; jinak budeme dekódovat $\mathbf{y}$ jako libovolné jiné kódové slovo, např. $\mathbf{c}_1$.

Začněme nyní s vlastním důkazem. Nechť $\mathbf{Y}$ je vektor, který obdržíme, když je přenášeno kódové slovo $\mathbf{c}$ a E buď událost, že nastala chyba.

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Pak, je-li $\mathbf{y}$ obdržený vektor, můžeme dekódovat $\mathbf{y}$ jako kódové slovo $\mathbf{c}_j$, je-li $\mathbf{c}_j$ jediné kódové slovo v $S_r(\mathbf{y})$; jinak budeme dekódovat $\mathbf{y}$ jako libovolné jiné kódové slovo, např. $\mathbf{c}_1$.

Začněme nyní s vlastním důkazem. Nechť $\mathbf{Y}$ je vektor, který obdržíme, když je přenášeno kódové slovo $\mathbf{c}$ a E buď událost, že nastala chyba.

Přitom chyba může nastat právě tehdy, když buď

(a) $$d(\mathbf{c}, \mathbf{Y}) > r$$

nebo

(b) $d(\mathbf{c}, \mathbf{Y}) \leq r$ a $d(\mathbf{c'}, \mathbf{Y}) \leq r$ pro nějaké jiné kódové slovo $\mathbf{c'}$. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Označme po řadě A a B události popsané (a) a (b). Pak $E = A \cup B$ a tudíž

$$P(E) = P(A \cup B) \leq P(A) + P(B).$$

Uvažme událost B. Ta nastane, pokud platí zároveň

(i) Ne více než r chyb nastalo při přenosu,

(ii) jedno z kódových slov různých od $\mathbf{c}$ je ve vzdálenosti nejvýše $\mathbf{r}$ od obdrženého vektoru $\mathbf{Y}$. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Označíme-li po řadě tyto události B_1 a B_2, máme pak, protože $B = B_1 \cap B_2$,

$$P(B) \leq P(B_2). \tag{6.3}$$

Uvažme nyní B_2; protože kódová slova jsou vybrána náhodně, pravděpodobnost, že $\mathbf{c}_i$ má vzdálenost menší nebo rovnu r od $\mathbf{Y}$ je $N_r(n)/2^n$, kde

$$N_r(n) = \sum_{k=0}^{r} \binom{n}{k} \tag{6.4}$$

je počet vektorů z V_n, které leží v $S_r(\mathbf{y})$. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Tudíž pravděpodobnost, že alespoň jedno ze zbývajících $M-1$ kódových slov (různých od $\mathbf{c}$) má vzdálenost menší nebo rovnu r od obdrženého slova $\mathbf{Y}$ splňuje

$$P(B_2) \leq \frac{M-1}{2^n} \sum_{k=0}^{r} \binom{n}{k}. \tag{6.5}$$

Položme tudíž, pro všechna $\varepsilon > 0$,

$$r = \lfloor np + n\varepsilon \rfloor$$

jakožto maximální celé číslo ne větší než $np + n\varepsilon$. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Obdržíme pak z 6.3, 6.4, 6.5 a omezené nerovnosti, že

$$P(B) \leq \frac{M}{2^n} 2^{nh(p+\varepsilon)} = M2^{-n[1-h(p+\varepsilon)]}. \tag{6.6}$$

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Obdržíme pak z 6.3, 6.4, 6.5 a omezené nerovnosti, že

$$P(B) \leq \frac{M}{2^n} 2^{nh(p+\varepsilon)} = M2^{-n[1-h(p+\varepsilon)]}. \tag{6.6}$$

Věnujme se nyní druhému typu chyb způsobenému jevem A. Poznamenejme, že, je-li U (náhodný) počet chybných symbolů vzniklých při přenosu kódového slova $\mathbf{c}$, pak máme

$$P(A) = P(U > r)$$

a U je náhodná proměnná s binomiálním rozdělením s parametry n a p. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Tudíž

$$\begin{aligned} P(A) = P(U > np + n\varepsilon) &\leq P(|U - np| > n\varepsilon) \\ &\leq D(U)/n^2\varepsilon^2, \end{aligned}$$

dle Čebyševovy nerovnosti.

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Tudíž

$$\begin{aligned} P(A) = P(U > np + n\varepsilon) &\leq P(|U - np| > n\varepsilon) \\ &\leq D(U)/n^2\varepsilon^2, \end{aligned}$$

dle Čebyševovy nerovnosti.

Protože U je náhodná proměnná s binomiálním rozdělením, máme

$$D(U) = npq$$

a tedy úplná pravděpodobnost chyby je

$$P(E) \leq \frac{pq}{n\varepsilon^2} + M2^{-n[1-h(p+\varepsilon)]}.$$

pro dostatečně velká n. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Protože kapacita $C(p+\varepsilon) = 1 - h(p+\varepsilon)$, máme pak

$$P(E) \leq \frac{pq}{n\varepsilon^2} + M2^{-nC(p+\varepsilon)}.$$

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Protože kapacita $C(p+\varepsilon) = 1 - h(p+\varepsilon)$, máme pak

$$P(E) \leq \frac{pq}{n\varepsilon^2} + M2^{-nC(p+\varepsilon)}.$$

Protože $\varepsilon > 0$, lze pravděpodobnost chyby zvolit libovolně malou pro dostatečně velké n, za předpokladu, že M jakožto funkce n, neroste rychleji než $2^{nC(p)}$.

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Protože kapacita $C(p+\varepsilon) = 1 - h(p+\varepsilon)$, máme pak

$$P(E) \leq \frac{pq}{n\varepsilon^2} + M2^{-nC(p+\varepsilon)}.$$

Protože $\varepsilon > 0$, lze pravděpodobnost chyby zvolit libovolně malou pro dostatečně velké n, za předpokladu, že M jakožto funkce n, neroste rychleji než $2^{nC(p)}$.

Dokázali jsme tedy větu o kódování se šumem až na to, že jsme omezili průměrnou pravděpodobnost chyby a ne maximální pravděpodobnost chyby.
K dokončení důkazu potřebujeme dokázat, že existují kódy $\mathcal{C}_n$ s M_n kódovými slovy, kde $M_n \leq 2^{Rn}$ a mající maximální pravděpodobnost chyby $< \varepsilon$. ∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Položme proto $\varepsilon' = \frac{1}{2}\varepsilon$ a $M'_n = 2M_n$.
Poznamenejme, že protože $M_n \leq 2^{Rn}$ a $R < C$, musí existovat R' tak, že $R < R' < C$, a N'_0 tak, že pro všechna $n \geq N'_0$ platí

$$M'_n \leq 2^{nR'}$$

a tudíž existuje posloupnost kódů $\mathcal{C}'_n$ tak, že $\mathcal{C}'_n$ má M'_n kódových slov a ***průměrnou*** pravděpodobnost chyby $< \varepsilon'$ pro $n \geq N'_0$.
Jsou-li $\mathbf{x}_1, \ldots, \mathbf{x}_{M'_n}$ kódová slova z $\mathcal{C}'_n$, znamená to, že

$$\frac{1}{M'_n} \sum_{i=1}^{M'_n} P(E|\mathbf{x}_i) = \sum_{i=1}^{M'_n} P(E|\mathbf{x}_i) \cdot P(\mathbf{x}_i) = P(E) \leq \varepsilon'.$$

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Tedy alespoň polovina těchto kódových slov $\mathbf{x}_i$ musí splňovat

$$P(E|\mathbf{x}_i) \leq 2\varepsilon' = \varepsilon. \tag{6.7}$$

Buď $\mathcal{C}_n$ kód sestávající z M_n kódových slov splňujících 6.7; pak máme náš požadovaný kód s maximální pravděpodobností $\leq \varepsilon$.

∎

Pokračování důkazu věty o kódování se šumem

Důkaz Věty 6.1.

Tedy alespoň polovina těchto kódových slov $\mathbf{x}_i$ musí splňovat

$$P(E|\mathbf{x}_i) \leq 2\varepsilon' = \varepsilon. \tag{6.7}$$

Buď $\mathcal{C}_n$ kód sestávající z M_n kódových slov splňujících 6.7; pak máme náš požadovaný kód s maximální pravděpodobností $\leq \varepsilon$.
∎

Shannonovu větu lze rozšířit i pro obecné kanály bez paměti s libovolnou vstupní a výstupní abecedou. Hlavní myšlenka důkazu se nemění, totiž
(a) zakódujme zprávy náhodně,
(a) dekódujme procedurou maximální pravděpodobnosti.

Zesílení věty o kódování se šumem

Technické obtíže jsou způsobeny zejména obecným tvarem kapacity kanálu, pokud se nejedná o symetrický kanál. Případný zájemce může najít úplný důkaz (ve skutečnosti dva) pro tuto obecnou situaci v článku Ashe (1965) nebo Gallagera (1968).

Zesílení věty o kódování se šumem

Technické obtíže jsou způsobeny zejména obecným tvarem kapacity kanálu, pokud se nejedná o symetrický kanál. Případný zájemce může najít úplný důkaz (ve skutečnosti dva) pro tuto obecnou situaci v článku Ashe (1965) nebo Gallagera (1968).

Měli bychom se též zmínit o důležitosti zlepšení hranic pravděpodobnosti vzniku chyby. V našem důkazu nahoře nás pouze zajímalo to, že pravděpodobnost nastání chyby lze dosáhnout libovolně malou. K tomuto problému existuje bohatá a dostatečně technická literatura.

Zesílení věty o kódování se šumem

Například následující silnější výsledek uvedený bez důkazu přináleží Shannonovi (1957).

Věta 6.2

Buď dán diskrétní kanál bez paměti kapacity C a libovolné R, $0 < R < C$. Pak existuje posloupnost kódů $(\mathcal{C}_n : 1 \leq n < \infty)$ tak, že:
(a) $\mathcal{C}_n$ má $\lfloor 2^{Rn} \rfloor$ kódových slov délky n
(b) maximální pravděpodobnost chyby $\hat{e}(\mathcal{C}_n)$ kódování $\mathcal{C}_n$ splňuje

$$\hat{e}(\mathcal{C}_n) \leq Ae^{-Bn},$$

přičemž A a B závisí pouze na kanálu a na R.

Jinak řečeno, neexistují pouze dobré kódy, ale navíc existují kódy, jejichž pravděpodobnost chyby klesá exponenciálně.

Věta o kódování se šumem - Cvičení

Cvičení 6.3

1. *Binární symetrický kanál mající pravděpodobnost chyby přenosu $p = 0.05$ může přenést 800 binárních číslic za sekundu. Kolik bitů může přenést bez chyby za sekundu?*
2. *Binární symetrický kanál s fyzikální kapacitou přenosu 800 číslic za sekundu může přenést 500 číslic za sekundu s libovolně malou pravděpodobností chyby. Co nám to vypovídá o pravděpodobnosti chyby tohoto kanálu?*

Obsah

Kapacita jako hranice spolehlivé komunikace

FI Předpokládejme, že máme diskrétní kanál bez paměti o kapacitě C bitů. Předpokládejme, že tento kanál má mechanickou rychlost jednoho bitu za sekundu.

Dokážeme nyní obrácení Shannonovy věty tím, že ukážeme nemožnost přenosu přesné informace rychlostí vyšší nebo rovné než je C bitů za sekundu.

Přesněji, dokážeme následující základní výsledek.

Kapacita jako hranice spolehlivé komunikace

Věta 7.1

Pro kanál bez paměti o kapacitě C a pro každé $R > C$ neexistuje posloupnost kódů $(\mathcal{C}_n : 1 \leq n < \infty)$ tak, že:

(a) *$\mathcal{C}_n$ má 2^{Rn} kódových slov délky n,*

(b) *pravděpodobnost chyby $e(\mathcal{C}_n)$ kódování $\mathcal{C}_n$ konverguje k nule pro $n \to \infty$.*

Kapacita jako hranice spolehlivé komunikace

Věta 7.1

Pro kanál bez paměti o kapacitě C a pro každé $R > C$ neexistuje posloupnost kódů $(\mathcal{C}_n : 1 \leq n < \infty)$ tak, že:

(a) *$\mathcal{C}_n$ má 2^{Rn} kódových slov délky n,*

(b) *pravděpodobnost chyby $e(\mathcal{C}_n)$ kódování $\mathcal{C}_n$ konverguje k nule pro $n \to \infty$.*

Ve skutečnosti Wolfowitz v roce 1961 dokázal mnohem silnější výsledek – totiž, za stejných podmínek, maximální pravděpodobnost chyby konverguje k 1 pro $n \to \infty$. My však ukážeme slabší verzi, abychom dokázali, že Shannonova věta je nejlepší možná. Pro důkaz věty potřebujeme následující lemmata.

Kapacita jako hranice spolehlivé komunikace

Lemma 7.2

Buď $\mathbf{U}, \mathbf{V}, \mathbf{W}$ *náhodné vektory. Pak platí*

$$H(\mathbf{U}|\mathbf{V}) \leq H(\mathbf{U}|\mathbf{V},\mathbf{W}) + H(\mathbf{W}).$$

Důkaz.

Máme dle základní identity, že

$$\begin{aligned} H(\mathbf{U}|\mathbf{V}) &= H(\mathbf{U},\mathbf{V}) - H(\mathbf{V}) \\ &= H(\mathbf{U},\mathbf{V},\mathbf{W}) - H(\mathbf{W}|\mathbf{U},\mathbf{V}) - H(\mathbf{V}) \\ &\leq H(\mathbf{U},\mathbf{W}|\mathbf{V}), \end{aligned}$$

protože entropie je nezáporná. ∎

Kapacita jako hranice spolehlivé komunikace

Pokračování důkazu Lemmatu 7.2.

Ale zároveň

$$\begin{aligned} H(\mathbf{U},\mathbf{W}|\mathbf{V}) &= H(\mathbf{U},\mathbf{V},\mathbf{W}) - H(\mathbf{V},\mathbf{W}) + H(\mathbf{V},\mathbf{W}) - H(\mathbf{V}) \\ &= H(\mathbf{U}|\mathbf{V},\mathbf{W}) + H(\mathbf{W}|\mathbf{V}) \\ &\leq H(\mathbf{U}|\mathbf{V},\mathbf{W}) + H(\mathbf{W}), \end{aligned}$$

což se mělo dokázat. ∎

Fanova nerovnost

Lemma 7.3

Fanova nerovnost *Buď C kód s M kódovými slovy $\{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$ pro daný kanál bez paměti.*
Buď $\mathbf{X}$ náhodný vektor nabývající hodnoty v množině kódových slov.
Nechť $\mathbf{Y}$ obsahuje náhodný vektorový výstup, v případě, že $\mathbf{X}$ je přeneseno kanálem a dekódováno.
Pak, je-li p_E pravděpodobnost chyby (totiž $p_E = P(\mathbf{X} \neq \mathbf{Y})$), máme

$$H(\mathbf{X}|\mathbf{Y}) \leq H(p_E, q_E) + p_E \log(M-1), \qquad (7.1)$$

kde $q_E = 1 - p_E$.

Fanova nerovnost

Důkaz Lemmatu 7.3.

Definujme novou náhodnou proměnnou Z jakožto

$$Z = \begin{cases} 0 & \text{pokud } \mathbf{X} = \mathbf{Y} \\ 1 & \text{pokud } \mathbf{X} \neq \mathbf{Y}. \end{cases}$$

Je tedy speciálně entropie náhodné proměnné Z rovna $H(p_E, q_E)$. Uvažme nyní uspořádanou dvojici $(\mathbf{Y}, Z)$. Zřejmě pak

$$H(\mathbf{X}|(\mathbf{Y}, Z) = (\mathbf{y}, 0)) = 0.$$

Zároveň, pokud $(\mathbf{Y}, Z) = (\mathbf{y}, 1)$, je náhodná proměnná $\mathbf{X}$ rozložena mezi $(M - 1)$ kódovými slovy, která nejsou rovna $\mathbf{y}$.

∎

Fanova nerovnost

Pokračování důkazu Lemmatu 7.3.

Zejména tedy

$$H(\mathbf{X}|(\mathbf{Y},Z)=(\mathbf{y},1)) \leq \log_2(M-1).$$

a

$$\begin{aligned} H(\mathbf{X}|(\mathbf{Y},Z)) &= \textstyle\sum_{\mathbf{y}} H(\mathbf{X}|(\mathbf{Y},Z)=(\mathbf{y},1)) \cdot P((\mathbf{Y},Z)=(\mathbf{y},1)) \\ &\leq \log_2(M-1) \textstyle\sum_{\mathbf{y}} \cdot P((\mathbf{Y},Z)=(\mathbf{y},1)) \\ &\leq p_E \cdot \log_2(M-1). \end{aligned}$$

Položme pak $\mathbf{U}=\mathbf{X}$, $\mathbf{V}=\mathbf{Y}$ a $\mathbf{W}=Z$. Z lemmatu 7.2 máme Fanovu nerovnost.

∎

Důkaz Věty 7.1

Předpokládejme, že takováto posloupnost kódů existuje. Uvažme pak náhodný vektor $\mathbf{X}$, který nabývá hodnot v kódu $\mathcal{C}_n$ tak, že pokud položíme $R = C + \varepsilon$, $\varepsilon > 0$, máme

$$H(\mathbf{X}) = n(C + \varepsilon).$$

Totiž $|\mathcal{C}_n| = 2^{R \cdot n}$ a vždy jde najít n-rozměrný náhodný vektor $\mathbf{X}$ s příslušným rovnoměrným rozdělením pravděbodnosti. Protože kapacita kanálu je C, máme pak pro kódová slova délky n, že odpovídající kapacita rozšíření bez paměti je nC a tedy, označíme-li $\mathbf{Y}$ náhodný vektor výstupu odpovídající vstupnímu náhodnému vektoru $\mathbf{X}$, máme nerovnost

$$H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \leq nC,$$

takže

$$n\varepsilon = n(C + \varepsilon) - nC \leq H(\mathbf{X}|\mathbf{Y}).$$

Důkaz Věty 7.1 - pokračování

Aplikujeme-li Fanovu nerovnost, pak z toho, že máme dle předpokladu $2^{n(C+\varepsilon)}$ kódových slov, je

$$n\varepsilon \leq H(\mathbf{X}|\mathbf{Y}) \leq H(p_E, q_E) + p_E \log(M-1) \leq H(p_E, q_E) + p_E n(C+\varepsilon),$$

tj.

$$\frac{n\varepsilon - H(p_E, q_E)}{n(C+\varepsilon)} \leq p_E.$$

Necháme-li n konvergovat k nekonečnu, pak zcela jistě p_E nekonverguje k nule. Tedy takováto posloupnost kódů $\mathcal{C}_n$ nemůže existovat.

Obsah

Markovův řetězec I

Nerovnost při zpracování dat může být použita k prokázání, že žádná chytrá manipulace s daty nemůže zlepšit závěry, které lze získat z dat.

Markovův řetězec I

Nerovnost při zpracování dat může být použita k prokázání, že žádná chytrá manipulace s daty nemůže zlepšit závěry, které lze získat z dat.

Řekneme, že ***náhodné proměnné X, Y a Z tvoří Markovův řetězec*** v tomto pořadí (označený $X \to Y \to Z$), pokud podmíněné rozdělení Z závisí pouze na Y a je podmíněně nezávislé na X. Přesněji, X, Y a Z tvoří Markovův řetězec $X \to Y \to Z$, pokud lze zapsat sdruženou pravděpodobnostní funkci jakožto

$$p(x, y, z) = p(x)p(y|x)p(z|y).$$

Markovův řetězec II

Lemma 8.1

(i) *$X \to Y \to Z$ právě tehdy, když X a Z jsou podmínečně nezávislé za podmínky Y.*

Markovův řetězec II

Lemma 8.1

(i) $X \to Y \to Z$ *právě tehdy, když* X *a* Z *jsou podmínečně nezávislé za podmínky* Y.

(ii) $X \to Y \to Z$ *právě tehdy, když* $Z \to Y \to X$.

Markovův řetězec II

Lemma 8.1

(i) $X \to Y \to Z$ *právě tehdy, když* X *a* Z *jsou podmínečně nezávislé za podmínky* Y.

(ii) $X \to Y \to Z$ *právě tehdy, když* $Z \to Y \to X$.

(iii) *Pokud* $Z = f(Y)$, *pak* $X \to Y \to Z$.

Markovův řetězec II

Lemma 8.1

(i) $X \to Y \to Z$ *právě tehdy, když* X *a* Z *jsou podmínečně nezávislé za podmínky* Y.

(ii) $X \to Y \to Z$ *právě tehdy, když* $Z \to Y \to X$.

(iii) *Pokud* $Z = f(Y)$, *pak* $X \to Y \to Z$.

Důkaz.

(i) Nechť $X \to Y \to Z$. Pak

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

Obráceně máme $p(x,y) = p(x)p(y|x)$ a

$$p(x,y,z) = p(x,z|y)p(y) = p(x|y)p(z|y)p(y) = p(x,y)p(z|y).$$

Markovův řetězec III

Důkaz.

(ii) Plyne bezprostředně z (i), protože máme

$$p(z, x|y) = p(z|y)p(x|y).$$

(iii) Je zřejmé, protože $p(z|y) = \begin{cases} 1 & \text{pokud } z = f(y) \\ 0 & \text{jinak.} \end{cases}$ ∎

Markovův řetězec III

Důkaz.

(ii) Plyne bezprostředně z (i), protože máme

$$p(z, x|y) = p(z|y)p(x|y).$$

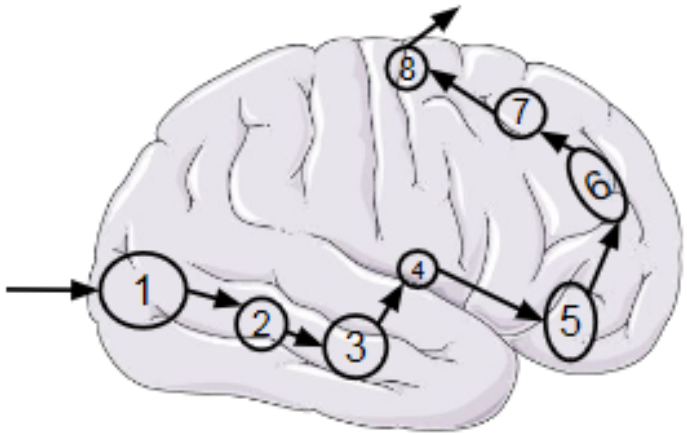
(iii) Je zřejmé, protože $p(z|y) = \begin{cases} 1 & \text{pokud } z = f(y) \\ 0 & \text{jinak.} \end{cases}$ ∎

Nyní dokážeme důležitou a užitečnou větu, která demonstruje, že žádné zpracování Y, deterministické nebo náhodné, nemůže zvýšit informaci, kterou Y obsahuje o X.

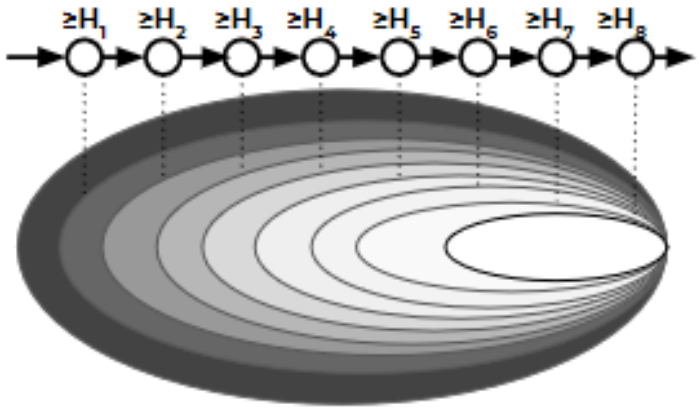
Markovův řetězec IV

Data processing inequality

$$\forall g,\ H(u) \geq H(g(u))$$



Unfolded



Nerovnost při zpracování dat I

Lemma 8.2

Buď U, V, W náhodné proměnné. Pak

(i) $I(U|V) = I(V|U) = H(U,V) - H(U|V) - H(V|U)$,

Nerovnost při zpracování dat I

Lemma 8.2

Buď U, V, W náhodné proměnné. Pak

(i) $I(U|V) = I(V|U) = H(U,V) - H(U|V) - H(V|U)$,

(ii) $H(U,V|W) = H(V|W) + H(U|V,W)$.

Nerovnost při zpracování dat I

Lemma 8.2

Buď U, V, W náhodné proměnné. Pak

(i) $I(U|V) = I(V|U) = H(U,V) - H(U|V) - H(V|U)$,

(ii) $H(U,V|W) = H(V|W) + H(U|V,W)$.

Důkaz.

$$\begin{aligned} H(U,V) - H(U|V) - H(V|U) &= \\ &= H(U,V) - [H(U,V) - H(V)] - [H(U,V) - H(U)] \\ &= H(U) + H(V) - H(U,V) = I(U|V) = I(V|U), \\ H(V|W) + H(U|V,W) &= \\ &= [H(V,W) - H(W)] + [H(U,V,W) - H(V,W)] \\ &= H(U,V,W) - H(W) = H(U,V|W). \end{aligned}$$

Nerovnost při zpracování dat II

Lemma 8.3

Buď X, Y, Z náhodné proměnné. Pak následující podmínky jsou ekvivalentní:

(i) $X \to Y \to Z$,

Nerovnost při zpracování dat II

Lemma 8.3

Buď X, Y, Z náhodné proměnné. Pak následující podmínky jsou ekvivalentní:

(i) $X \to Y \to Z$,

(ii) $H(X, Z|Y = y) = H(X|Y = y) + H(Z|Y = y)$, *pokud* $P(Y = y) > 0$,

Nerovnost při zpracování dat II

Lemma 8.3

Buď X, Y, Z náhodné proměnné. Pak následující podmínky jsou ekvivalentní:

(i) $X \to Y \to Z$,

(ii) $H(X, Z|Y = y) = H(X|Y = y) + H(Z|Y = y)$, *pokud* $P(Y = y) > 0$,

(iii) $H(X, Z|Y) = H(X|Y) + H(Z|Y)$,

Nerovnost při zpracování dat II

Lemma 8.3

Buď X, Y, Z náhodné proměnné. Pak následující podmínky jsou ekvivalentní:

(i) $X \to Y \to Z$,

(ii) $H(X, Z|Y = y) = H(X|Y = y) + H(Z|Y = y)$, *pokud* $P(Y = y) > 0$,

(iii) $H(X, Z|Y) = H(X|Y) + H(Z|Y)$,

(iv) $H(X|Y, Z) = H(X|Y)$.

Nerovnost při zpracování dat III

Důkaz.

Protože $H(X,Z|Y) = \sum_{P(Y=y)>0} P(Y=y)H(X,Z|Y=y)$, $H(X|Y) = \sum_{P(Y=y)>0} P(Y=y)H(X|Y=y)$ a $H(Z|Y) = \sum_{P(Y=y)>0} P(Y=y)H(Z|Y=y)$, a zároveň $H(X,Z|Y=y) \leq H(X|Y=y) + H(Z|Y=y)$ pro $P(Y=y) > 0$, je nutně podmínka (ii) ekvivalentní s (iii).

Dále $H(X,Z|Y) = H(X|Y) + H(Z|Y)$ právě tehdy, když

$$H(X,Y,Z) - H(Y) = \Big[H(X,Y) - H(Y)\Big] + \Big[H(Z,Y) - H(Y)\Big]$$

právě tehdy, když $H(X,Y,Z) - H(Y,Z) = H(X,Y) - H(Y)$ právě tehdy, když $H(X|Y,Z) = H(X|Y)$.
Tedy podmínka (iv) je ekvivalentní s (iii).

∎

Nerovnost při zpracování dat IV

Nechť tedy platí (i) a $P(Y = y) > 0$. Pak jsou náhodné proměnné $X|Y = y$ a $Z|Y = y$ s pravděpodobnostními rozděleními $p(x|y)$ a $p(z|y)$ nezávislé a náhodný vektor $(X|Y = y, Z|Y = y)$ můžeme ztotožnit s náhodným vektorem $(X, Z|Y = y)$ mající rozdělení $p(x|y)p(z|y)$. Tedy nutně $H(X, Z|Y = y) = H(X|Y = y) + H(Z|Y = y)$.

Obráceně, nechť platí

$$H(X|Y = y, Z|Y = y) = H(X, Z|Y = y) = H(X|Y = y) + H(Z|Y = y).$$

Pak jsou náhodné proměnné $X|Y = y$ a $Z|Y = y$ nezávislé a tedy $p(z, x|y) = p(z|y)p(x|y)$, tj., $X \to Y \to Z$.

Nerovnost při zpracování dat V

Věta 8.4

(Nerovnost při zpracování dat) Pokud $X \to Y \to Z$, *pak* $I(X|Y) \geq I(X|Z)$.

Důkaz.

$$\begin{aligned} I(X|Z) &= \Big[\overbrace{H(X,Y,Z) - H(Y|X,Z)}^{H(X,Z)}\Big] - \Big[\overbrace{H(X,Y|Z) - H(Y|X,Z)}^{H(X|Z)}\Big] \\ &\quad - \Big[\underbrace{H(Y,Z|X) - H(Y|X,Z)}_{H(Z|X)}\Big] \\ &= H(X,Y,Z) + H(Y|X,Z) - H(X,Y|Z) - H(Y,Z|X). \end{aligned}$$

∎

Nerovnost při zpracování dat VI

$$
\begin{aligned}
I(X|Z) &= H(X,Y,Z) + H(Y|X,Z) - H(X,Y|Z) - H(Y,Z|X) \\
&= H(X,Y,Z) + H(Y|X,Z) - \Big[H(X|Y,Z) + H(Y|Z)\Big] \\
&\quad - \Big[H(Z|Y,X) + H(Y|X)\Big] \\
&= H(X,Y,Z) + H(Y|X,Z) - \Big[H(X|Y) + H(Y|Z)\Big] \\
&\quad - \Big[H(Z|Y) + H(Y|X)\Big] \\
&= \Big[H(X,Y) + H(X|Y)\Big] + H(Y|X,Z) \\
&\quad - \Big[H(X|Y) + H(Y|Z) + H(Z|Y) + H(Y|X)\Big] \\
&= \Big[H(X,Y) - H(X|Y) - H(Y|X)\Big] + \Big[H(Y|X,Z) - H(Y|Z)\Big] \\
&= I(X|Y) + \underbrace{H(Y|X,Z) - H(Y|Z)}_{\leq 0}
\end{aligned}
$$

Tedy $I(X|Z) \leq I(X|Y)$.

Nerovnost při zpracování dat VII

Důsledek 8.5

Pokud $X \to Y \to Z$, *pak*

(i) $I(Z|X) \leq I(Y|X)$,

(ii) $I(X|Z) \leq I(Y|Z)$,

(iii) *je-li* g *reálná funkce, pak* $I(g(Y)|X) \leq I(Y|X)$.

Důkaz.

(i) je reformulace Věty 8.4, (ii) obdržíme z toho, že máme $Z \to Y \to X$ a dosazením do (i). Poslední část plyne z toho, že $X \to Y \to g(Y$ je Markovův řetězec a Věty 8.4. ∎

Nerovnost při zpracování dat VIII

Na výklad věty lze nahlížet následujícím způsobem. Předpokládejme, že nejprve se X transformuje na Y procesem A. To může být například přenos dat přes kanál, který zkresluje signály (např. internetová komunikace nebo zápis a čtení CD, DVD, nebo flash paměti).

$$X \xrightarrow{A} Y \xrightarrow{B} Z$$

Získáme tak mnoho informací o X pozorováním Y. Dále je běžné provádět následné zpracování, které v tomto modelu představuje proces B.
Tvrzení věty potvrzuje, že informace o X zachycením Z nemohou překročit informace o X zachycením Y. Jinými slovy, informaci o X nelze zvětšit postprocessingem, může jen klesnout.

Nerovnost při zpracování dat IX

V praxi je však postprocessing často používán k transformaci informací do jiné reprezentace, kde jsou informace je snadněji přístupné pro interpretaci. Například je snazší pochopit obraz při prohlížení na obrazovce, než je tomu z přijatých dat.
Podobně může proces A představovat předzpracování a proces B přenos. Potom věta potvrzuje, že se informace nemohou zvyšovat předzpracováním.
Přesto je v praxi běžné používat předzpracování v komunikačních systémech pro transformaci dat do vhodných reprezentací.
Když to shrneme, věta tvrdí, že informace se nemohou zvětšovat ani předzpracováním ani následným zpracováním.
Informace se mohou během zpracování pouze snižovat.

Problémy

Problémy 1

1. *V binárním symetrickém kanálu s pravděpodobností chyby* $\epsilon > 0$, *kódování sestává ze dvou kódových slov* 000 *a* 111. *Zjistěte při použití pravidla maximální pravděpodobnosti pravděpodobnost chyby.*
2. Trhlinová chyba *(burst error) délky k sestává z posloupnosti k symbolů, které byly všechny přeneseny nesprávně. Najděte očekávaný počet trhlinových chyb délky k, pokud je zpráva délky N přenesena binárním symetrickým kanálem s pravděpodobností chyby p.*

Problémy k řešení

Problémy 1

3. *Nechť kód pro přenos binárním symetrickým kanálem, který má pravděpodobnost chyby* $\varepsilon > 0$, *sestává ze všech pětic nad množinou* $\{0, 1\}$, *které obsahují právě dvě jedničky. Jaká je pravděpodobnost, že kódové slovo* 11000 *se dekóduje na slovo* 10001, *pokud aplikujeme pravidlo minimální vzdálenosti?*
4. *Mějme N binárních symetrických kanálů, každý s pravděpodobností chyby p, spojených do série. Ukažte, že celková kapacita tohoto nově vzniklého kanálu je určena vztahem*

$$C_N = 1 + p_N \log p_N + q_N \log q_N,$$

kde $p_N = \frac{1}{2}[1 - (q - p)^N]$, $q_N = 1 - p_N$.

Problémy k řešení

Problémy 1

5. *Uvažme dva diskrétní kanály bez paměti o kapacitách C_1 a C_2 tak, že oba mají vstupní abecedu Σ_1 a výstupní abecedu Σ_2.* Součinem kanálů *je kanál, jehož vstupní abeceda je $\Sigma_1^{(2)}$ a výstupní abeceda $\Sigma_2^{(2)}$, přičemž kanálové pravděpodobnosti jsou určeny vztahem*

$$p(y_1 y_2|x_1 x_2) = p_1(y_1|x_1)p_2(y_2|x_2),$$

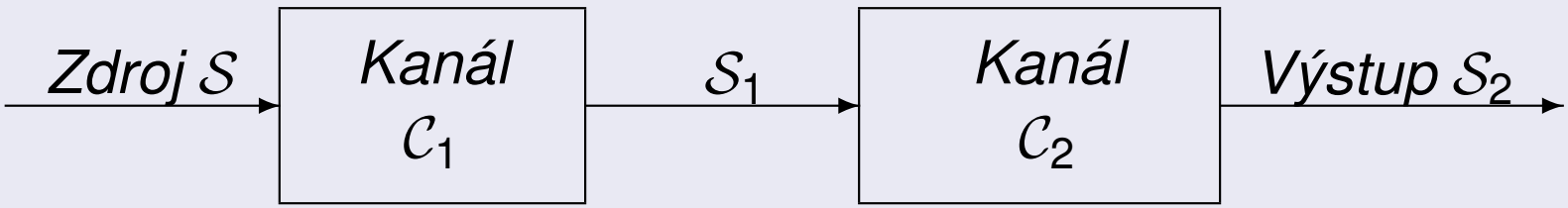
kde $p_i(y_i|x_i)$ je pravděpodobnost, že jsme obdželi řetězec y_i, pokud jsme odeslali řetězec x_i prostřednictvím i-tého kanálu. Dokažte, že kapacita C součinu kanálů je určena vztahem (Shannon 1957)

$$C = C_1 + C_2.$$

Problémy k řešení

Problémy 1

6. *Zdroj bez paměti $\mathcal{S}$ je spojen ke kanálu $\mathcal{C}_1$ o kapacitě C_1 a výsledný výstup $\mathcal{S}_1$ je vstup ke kanálu $\mathcal{C}_2$ o kapacitě C_2 (viz níže uvedený diagram).*



Obrázek 4: *Blokový diagram sdělovacího systému Příkladu 6.*

Ukažte, že platí

$$I(\mathcal{S}|\mathcal{S}_2) \leq I(\mathcal{S}|\mathcal{S}_1) \quad \text{a} \quad I(\mathcal{S}|\mathcal{S}_2) \leq I(\mathcal{S}_1|\mathcal{S}_2).$$