MAP a ML dekódování

Předpokládejme, že jsme při přenosu kódové zprávy kanálem obdrželi výstup r. Na základě tohoto výstupu chceme určit, jaká kódová zpráva v byla vyslána. Protože hodnota r je závislá jak na kódové zprávě, tak na šumu kanálu, nemůžeme v určit s jistotou. Zajímá nás ale, jaká kódová zpráva byla na záladě dostupné informace vyslána s největší pravděpodobností. Výstup r tedy opravíme na předpokládanou kódovou zprávu $\widehat{v}$, která je definována jako

$$\widehat{v} = \operatorname*{argmax}_v \Pr[v \mid r] .$$

Tento postup se nazývá MAP-dekódování, z anglického *Maximum Aposteriori Probability* („největší aposteriorní pravděpodobnost", či snad „největší pravděpodobnost podle dostupných výsledků").

MAP-dekódování nemůže být zaměňováno s otázkou, které z kódových slov je výstupu „nejpodobnější", či přesněji, pro kterou kódovou zprávu je nejpravděpodobnější, že se změní na daný výstup. Taková otázka se ptá na

$$\operatorname*{argmax}_v \Pr[r \mid v] ,$$

a volba tohoto maxima se nazývá ML-dekódování z anglického *Maximum Likelihood* („největší (pravdě)podobnost"). Mezi oběma postupy platí na základě rovností

$$\Pr[v, r] = \Pr[r \mid v] \Pr[v] = \Pr[v \mid r] \Pr[r]$$

vztah

$$\operatorname*{argmax}_v \Pr[v \mid r] = \operatorname*{argmax}_v \frac{\Pr[r \mid v] \Pr[v]}{\Pr[r]} = \operatorname*{argmax}_v \Pr[r \mid v] \Pr[v] .$$

Vidíme tedy, že výsledky MAP-dekódování a ML-dekódování mohou být různé, pokud nejsou všechny kódové zprávy stejně pravděpodobné. To vysvětluje význam často přijímaného předpokladu, že zpráva se vybírá uniformně náhodně.

Poznámka: Pravděpodobnostní prostor Ω implicitně přítomný ve výše uvedených úvahách je prostor všech událostí „přenos zprávy". Taková událost zahrnuje výběr kódové zprávy (tedy obvykle výběr informační zprávy a její zakódování) a její přenos kanálem (tedy všechny události ovlivňující hodnota výstupu). Náhodná veličina V přiřazuje každé takové události hodnotu kódové zprávy a náhodná veličina R hodnotu výstupu. Formule

$$\Pr[v \mid r]$$

je pak zkratkou za $\Pr[V = v \mid R = r]$.

*

Příklad: Rozdíl mezi MAP a ML dekódováním lze ilustrovat na příkladu binárního symetrického kanálu s chybovostí 0.1, po kterém jsou přenášeny jednotlivé zprávy $\{a, b\}$ zakódované takto: $a \mapsto 0$, $b \mapsto 1$. Zpráva b nechť má přitom frekvenci vyslání pouze 0.01 (můžeme to např. chápat jako hlášení o nějaké vzácně se vyskytující poruše nebo katastrofě). Při ML-dekódování dékodujeme $0 \mapsto a$, $1 \mapsto b$, protože pravděpodobnost chyby je menší než polovina. Při MAP-dekódování je ovšem při přijetí zprávy 1 pravděpodobnost $0.01 \cdot 0.9 = 0.009$, že bylo zamýšlenou zprávou skutečně b (a nedošlo k chybě), a pravděpodobnost $0.99 \cdot 0.1 = 0.099$, že byla vyslána nula (jako obvykle) a k chybě došlo. Je tedy jedenáctkrát pravděpodobnější, že zpráva je a.

Z toho je vidět, že takovýto přenos informace je bezcenný. Každý symbol nese zhruba 0.08 bitu informace. Šum kanálu je ovšem téměř 0.45 bitu. Většina informace, kterou dostáváme, je tedy šum. Současně ale platí, že kanál má kapacitu zhruba 0.55 bitu na symbol. Jeden bit informace by tedy mělo být (průměrně!) možné přenést dvěma symboly. Slovo „průměrně“ je důležité, protože jistě nelze přenést jeden bit informace pomocí bitů dvou. Dvě po sobě následující jedničky (jako kód pro zprávu b) se stále ještě s pravděpodobností $0.99 \cdot (0.1)^2 = 0.0099$ proti $0.01 \cdot (0.9)^2 = 0.0081$ dekóduje na a. Shannonova věta pouze garantuje, že pro libovolně malou toleranci chyby ε existuje n a nějaké kódování takové, že pomocí n symbolů přeneseme s požadovanou jistotou alespoň $0.55n$ bitů informace.