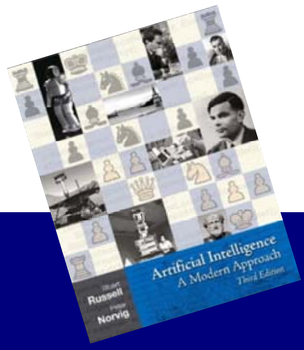
Umělá inteligence II



6

Roman Barták, KTIML

roman.bartak@mff.cuni.cz
http://ktiml.mff.cuni.cz/~bartak

Racionální rozhodování

- Připomeňme, že naším cílem je tvorba **racionálních agentů** maximalizujících očekávanou míru užitku.
- **Teorie pravděpodobnosti** nám dává nástroj, pro zjištění stavu světa případně efektu akce.
- Dnes se podíváme na to, jak ohodnotit rozhodnutí pomocí **teorie užitku**
  - a jak nakonec vybrat akci s největším očekávaným užitkem (**teorie rozhodování**).

Teorie užitku

- Agentovy preference lze zachytit **funkcí užitku** U, která mapuje stavy na reálné čísla.
- **Očekávaný užitek** potom spočteme jako průměrný užitek přes všechny možné stavy

$$EU(a|\mathbf{e}) = \Sigma_s P(Result(a)=s|a,\mathbf{e})\ U(s)$$

- Racionální agent potom volí akci **maximalizující očekávaný užitek** (MEU)

$$action = argmax_a\ EU(a|\mathbf{e})$$

- MEU formalizuje racionalitu, ale jak budeme celý postup operačně realizovat?

Preference

- Agentovy **preference** se často vyjadřují relativním porovnáním:
  - A > B: agent preferuje A před B
  - A < B: agent preferuje B před A
  - A ~ B: agent mezi A a B nemá žádnou preferenci (nerozlišuje A a B)
- Co je A a B?
  - Mohou to být stavy světa, ale pro neurčité výstupy se používají loterie.
  - **Loterie** popisuje možné výstupy S_1, …, S_n, které se vyskytují s danými pravděpodobnostmi p_1, …, p_n.
    - $[p_1,S_1; …; p_n,S_n]$
- Příklad loterie (nabídka jídla v letadle) Chcete kuře nebo těstoviny?
  - [0.8,dobré kuře; 0.2,připečené kuře]
  - [0.7,dobré těstoviny; 0.3,rozvařené těstoviny]

Racionální preference
vlastnosti

- **Racionální preference** vedou k maximalizaci očekávaného zisku (užitku).
- Musí splňovat některé podmínky (pokud je nesplňují, lze ukázat, že vedou k vnitřní iracionalitě).
  - **uspořádanost**:
    $(A > B) \vee (A < B) \vee (A \sim B)$
  - **tranzitivita**:
    $(A < B) \wedge (B < C) \Rightarrow (A < C)$
  - **spojitost**:
    $(A > B > C) \Rightarrow \exists p\ [p,A;\ 1\text{-}p,C] \sim B$
  - **substituovatelnost**:
    $A \sim B \Rightarrow [p,A;\ 1\text{-}p,C] \sim [p,B;\ 1\text{-}p,C]$
  - **monotonie**:
    $A > B \Rightarrow (p > q \Leftrightarrow [p,A;\ 1\text{-}p,B] > [q,A;\ 1\text{-}q,B]$
  - **rozkladatelnost**:
    $[p,A;\ 1\text{-}p,\ [q,B;\ 1\text{-}q,C]] \sim [p,A;\ (1\text{-}p)q,B;\ (1\text{-}p)(1\text{-}q),C]$

Preference a užitek

- Axiomy racionálních preferencí nehovoří přímo o funkci užitku, ale lze najít funkci užitku popisující preference.
- **Existuje funkce užitku** vracející pro danou loterii reálné číslo tak, že:
  - $U(A) < U(B) \Leftrightarrow A < B$
  - $U(A) = U(B) \Leftrightarrow A \sim B$
- **Očekávaný užitek loterie** lze spočítat:
  - $U([p_1,S_1;\ \ldots;\ p_n,S_n]\ ) = \Sigma_i\ p_i\ U(S_i)$
- Takových funkcí užitku může existovat více
  - $U(S) = a\ U(S) + b$
- Racionální agent ani nemusí svoji funkci užitku znát, ale pozorováním jeho preferencí ji lze zrekonstruovat.

Funkce užitku

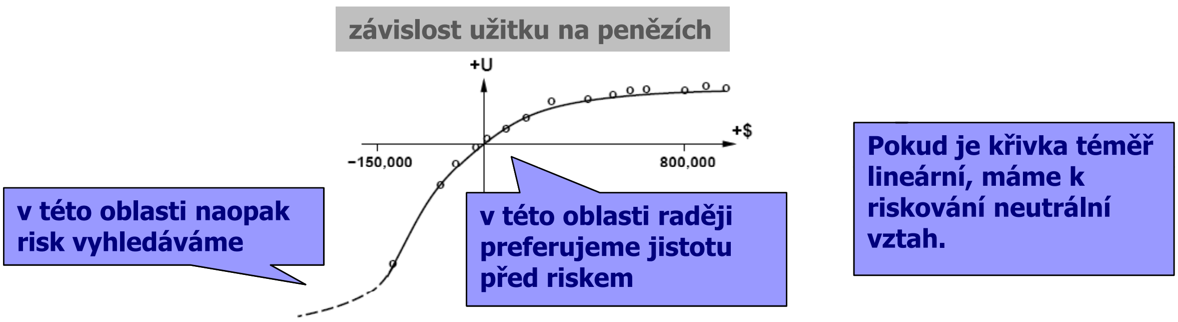
- Funkce užitku mapuje stavy/loterie na reálná čísla. Ale jaká konkrétní čísla?
- Jak zjistit funkci užitku konkrétního agenta (**preference eliciation**)?
  - Budeme hledat **normalizovanou funkci užitku**.
  - Uvažujme nejlepší možný stav S_{max} a dejme mu užitek 1, $U(S_{max}) = 1$.
  - Podobně pro nejhorší možný stav S_{min} dejme užitek 0, $U(S_{min}) = 0$.
  - Nyní se pro libovolný stav S ptejme agenta na porovnání S a standardní loterie $[p, S_{max}; 1\text{-}p, S_{min}]$
  - Podle výsledku upravíme p a ptáme se znova, dokud agent vztah A a standardní loterie nepovažuje za nerozlišitelný.
  - Získané p je užitkem S, $U(S) = p$.

Peníze jako užitek?

- V běžném životě používáme peníze pro ohodnocení různého zboží a služeb.
  - Agent zpravidla preferuje více peněz před méně penězi, je-li vše ostatní stejné.
- Proč nejsou peníze přímo mírou užitku?
- Uvažujme, že jsem vyhráli 1 mil. USD a můžeme si ho buď nechat nebo přijmeme sázku na hod mincí – padne-li orel dostaneme 2,5 mil. USD, jinak nic. Co zvolíte?
  - Očekávaný peněžní zisk při sázce je 1.250.000 USD.
  - Většina lidí ale volí jistotu 1 mil. USD. Je to snad iracionální?

Užitek z peněz

- Volba v předchozí hře závisí nejen na hře samé, ale i na současném majetku hráče!
- Nechť S_n je stav označující majetek n USD.
- Potom můžeme očekávaný užitek akcí popsat takto:
  - EU(Accept) = ½ U(S_k) + ½ U($S_{k+2.500.000}$)
  - EU(Decline) = U($S_{k+1.000.000}$)
- Nechť U(S_k) = 5, U($S_{k+1.000.000}$) = 8, U($S_{k+2.500.000}$) = 9.
- Potom je rozhodnutí odmítnout sázku zcela racionální!



Lidské rozhodování
efekt jistoty

- Lidé se chovají „předpovídatelně iracionálně“.
- **Allaisův paradox**
  - A: 80% šance dostat 4000 USD
  - B: 100% šance dostat 3000 USD

    Co zvolíte?
    - obvyklá volba je B, protože preferujeme jistotu
  - C: 20% šance dostat 4000 USD
  - D: 25% šance dostat 3000 USD

    Co zvolíte?
    - obvyklá volba je C, protože preferujeme větší očekávaný finanční zisk
- **Efekt jistoty** – lidé silně preferují zisk, který je jistý.

Lidské rozhodování
averze k nejednoznačnosti

- **Ellsbergův paradox**
  - V urně je 1/3 červených koulí a zbylé koule jsou černé nebo žluté.
  - A: pokud je vybrána červená koule, dostanete 100 USD.
  - B: pokud je vybrána černá koule, dostanete 100 USD.

    Co zvolíte?
    - obvyklá volba je A, možnost výhry je 1/3, zatímco pro B je možnost výhry mezi 0 a 2/3
  - C: 100 USD za vybrání červené nebo žluté koule
  - D: 100 USD za vybrání černé nebo žluté koule

    Co zvolíte?
    - obvyklá volba je D, možnost výhry je 2/3, zatímco pro C je možnost výhry mezi 1/3 a 3/3
  - Pokud si ale myslím, že v osudí je více červených než černých koulí, měl bych volit A a C
- **Averze k nejednoznačnosti** – lidé preferují známou nejistotu než nejistou nejistotu.

Lidské rozhodování
trochu psychologie

- **Efekt podání** – formulace problému má velký vliv na rozhodnutí agenta
  - Operace A dává 90% šanci na přežití
  - Operace B má 10% úmrtnost

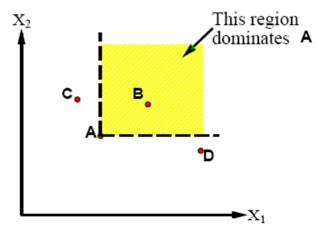
    Co zvolíte?
    - obvyklá volba je A, i když obě volby jsou naprosto totožné
- **Efekt kotvy** – lidé se cítí lépe při použití relativního porovnání než u absolutních hodnot
  - Proto v restauraci najdeme vína za 200 USD, která si stejně nikdo nekoupí (a restauratér to ví), ale víno za 50 USD potom vypadá jako dobrá koupě. Lidé totiž očekávají kvalitu všech vín podle toho nejdražšího.

Více atributů

- V praxi se často vyskytuje více atributů užitku například cena, nebezpečnost, užitečnost – **víceatributová funkce užitku**.
- Budeme uvažovat, že každý atribut má definované preferované pořadí hodnot, vyšší hodnoty odpovídají lepšímu řešení.
- Jak definovat preference pro více atributů dohromady?
  - přímo bez kombinace hodnot atributů – **dominance**
  - **kombinací hodnot** atributů do jedné hodnoty

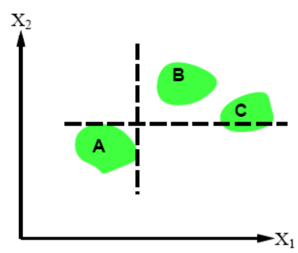
Dominance

- Pokud je ve všech atributech řešení A horší než řešení B, potom je B přirozeně lepší i celkově – **striktní dominance**.



Deterministic attributes

- Striktní dominanci lze definovat i pro **nejisté hodnoty** atributů.
  - stačí když každá možná hodnota všech atributů A je horší než každá možná hodnota odpovídajících atributů B



Uncertain attributes

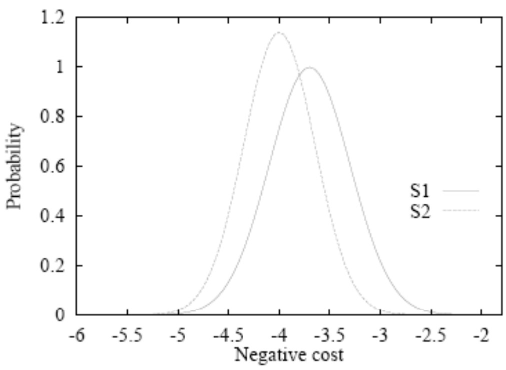
- Striktní dominance není moc obvyklá, ale může alespoň odfiltrovat „špatná“ řešení.

Stochastická dominance

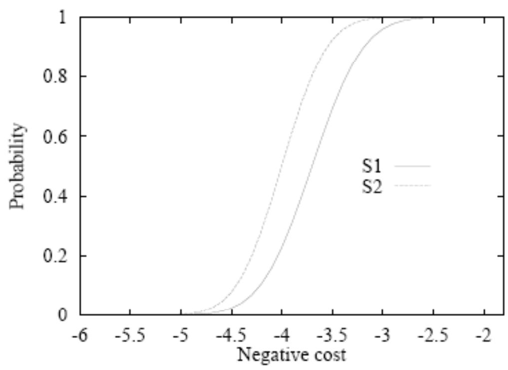
- Dominanci pro nejisté hodnoty můžeme definovat obecněji a tím umožnit více porovnání.
- **Stochastická dominace** je definována přes **kumulativní distribuci**, která měří pravděpodobnost, že cena je menší nebo rovna než daná hodnota.

$$\forall t \quad \int_{-\infty}^{t} p_1(x)dx \leq \int_{-\infty}^{t} p_2(t)dt$$

pravděpodobnostní distribuce



kumulativní distribuce



Struktura preferencí

- Pokud bychom chtěli popsat libovolnou **funkci užitku** pro n atributů, kde každý atribut má d hodnot, potřebovali bychom tabulku s **d^n hodnotami preferencí**.
  - To platí, pokud se mezi jednotlivými preferencemi nevyskytuje **žádná pravidelnost**.
- Často ale mají **preference jistou vnitřní strukturu** a funkci užitku lze popsat jako složení funkcí užitku jednotlivých atributů:

$$U(x_1,\dots,x_n) = F[f_1(x_1),\dots,f_n(x_n)]$$

Struktura preferencí
deterministické

- Základní pravidelnost mezi atributy je jejich **nezávislost**.
- Atributy X_1 a X_2 jsou **preferenčně nezávislé** na atributu X_3, pokud preferenční porovnání $\langle x_1,x_2,x_3\rangle$ a $\langle x`_1,x`_2,x_3\rangle$ nezáleží na konkrétní hodnotě x_3.
- Pokud je každá dvojice atributů preferenčně nezávislá na všech zbylých atributech, hovoříme o **vzájemné preferenční nezávislosti**.
- Máme-li množinu atributů s vzájemnou preferenční nezávislostí, potom

$$U(x_1,\ldots,x_n) = \Sigma_i\, U_i(x_i)$$

  - hovoříme o **aditivní oceňovací funkci**, může se jednat například o vážený součet.

Struktura preferencí
stochastické

- Nezávislost lze podobným způsobem definovat pro loterie (stochastické hodnoty).
- Pro skládání jednotlivých hodnot užitku se ale používá **multiplikativní funkce užitku**.
- Například pro složení tří atributů

$$\begin{aligned} U = {} & k_1U_1 + k_2U_2 + k_3U_3 \\ & + k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_1k_3U_1U_3 \\ & + k_1k_2k_3U_1U_2U_3 \end{aligned}$$

- Pro n atributů tedy potřebujeme n konstant, což je méně než exponenciální počet preferencí.

Hodnota informace

- Dosud jsme předpokládali, že máme k dispozici veškeré informace pro rozhodnutí.
- V praxi tomu tak ale často není, například lékař nemá k dispozici všechna možná vyšetření pacienta.
- **Jednou z nejdůležitějších součástí rozhodovacího procesu je vědět na co se ptát.**
- Teď se podíváme na **teorii hodnoty informace** (information value theory), která umožní agentovi rozhodnout, jaké informace je dobré získat.

Hodnota informace

jednoduchý příklad

- Uvažujme ropnou společnost, která může z n nerozlišitelných oblastí zakoupit jednu oblast pro těžbu.
- Dále uvažujme, že právě jedna oblast obsahuje ropu v ceně C dolarů. Cena nákupu libovolné oblasti je C/n.
  - Očekávaný zisk je tedy C/n – C/n = 0.
- Máme k dispozici seizmologa, který o jedné vybrané oblasti může s jistotou prohlásit, zda v ní je ropa nebo ne.
- **Kolik je cena takové informace?**
  - s pravděpodobností 1/n je v dané oblasti ropa, potom ji společnost koupí, takže zisk bude C – C/n
  - s pravděpodobností (n-1)/n tam ropa nebude, takže společnost koupí jinou oblast, kde je teď pravděpodobnost ropy 1/(n-1), očekávaný zisk je tedy C/(n-1) – C/n
  - dohromady je očekávaný zisk za předpokladu získání informace
    $1/n\ (C - C/n) + (n-1)/n\ (C/(n-1) - C/n) = C/n$

  Společnost je tedy ochotna zaplatit seizmologovi maximálně C/n, což je právě cena jedné oblasti.

Hodnota informace

obecný vzorec

- Předpokládejme, že známe hodnotu nějaké náhodné proměnné E_j – hovoříme potom o **hodnotě perfektní informace** (VPI-value of perfect information).
- Cena nejlepší akce α pro počáteční znalost

$$EU(\alpha|\mathbf{e}) = \max_a \Sigma_{s'} P(Result(a)=s'|a,\mathbf{e})\ U(s')$$

- Cena nejlepší akce α_{jk} po získání znalosti $E_j = e_{jk}$:

$$EU(\alpha_{jk}|\mathbf{e}, E_j=e_{jk}) = \max_a \Sigma_{s'} P(Result(a)=s'|a,\mathbf{e}, E_j=e_{jk})\ U(s')$$

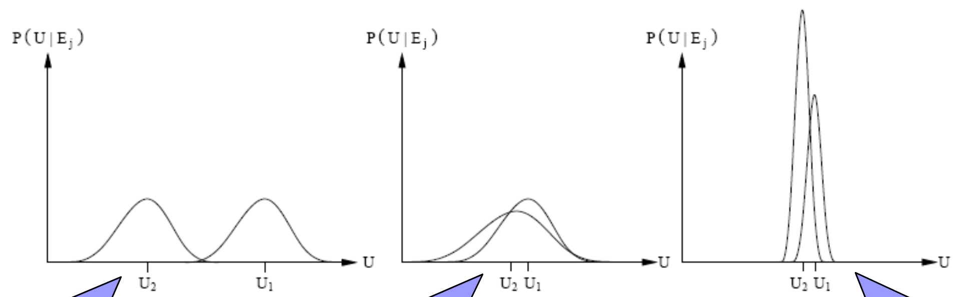
- My ale hodnotu proměnné E_j dopředu neznáme, pouze nás zajímá, jak by to vypadalo, pokud bychom ji znali.
- Vezmeme proto průměr přes všechny možné hodnoty E_j:

$$VPI_{\mathbf{e}}(E_j) = (\Sigma_k P(E_j = e_{jk}|\mathbf{e})\ EU(\alpha_{jk}|\mathbf{e}, E_j=e_{jk})) - EU(\alpha|\mathbf{e})$$

Hodnota informace

kvalitativní uvažování

- Kdy má cenu zjišťovat přesnou hodnotu některých atributů?



jasná volba, dodatečná informace je zbytečná

nejasná volba, dodatečná informace má velkou cenu

nejasná volba, dodatečná informace má ale malou cenu

- Informace má hodnotu do té míry, pokud
  - pravděpodobně vede ke změně plánu
  - nový plán je výrazně lepší než původně zamýšlený plán.

Hodnota informace

vlastnosti

- Může mít získaná informace negativní účinek?
- Očekávaná hodnota získané informace **není negativní**.

  $\forall \mathbf{e}, E_j \; VPI_{\mathbf{e}}(E_j) \geq 0$

- Očekávaná hodnota více získaných informací **není aditivní**.

  $VPI_{\mathbf{e}}(E_j,E_k) \neq VPI_{\mathbf{e}}(E_j) + VPI_{\mathbf{e}}(E_k)$

- Pokud máme více vstupních informací, **nezáleží** očekávaná hodnota **na pořadí**.

  $VPI_{\mathbf{e}}(E_j,E_k) = VPI_{\mathbf{e}}(E_j) + VPI_{\mathbf{e},ej}(E_k) = VPI_{\mathbf{e}}(E_k) + VPI_{\mathbf{e},ek}(E_j)$

Získávání informací

- Rozumný agent by měl
  - klást otázky v rozumném pořadí
  - vyhnout se irelevantním otázkám
  - vážit získanou informaci k ceně získání odpovědi
  - přestat klást otázky, když je to vhodné
- Uvažujme cenu získání hodnoty pozorované proměnné Ej, Cost(Ej).
- **Agent sbírající informace** může hladovým způsobem vybírat nejužitečnější proměnné, dokud „to má cenu".
(myopický/krátkozraký přístup)

```
function INFORMATION-GATHERING-AGENT(percept) returns an action
   static: D, a decision network

   integrate percept into D
   j ← the value that maximizes VPI(E_j) − Cost(E_j)
   if VPI(E_j) > Cost(E_j)
      then return REQUEST(E_j)
   else return the best action from D
```