1 Funkční závislosti stanovené z dat

Pro danou relaci $\mathcal{D}$ chceme najít, co možná nejmenší teorii T tak, že $\mathcal{D} \models A \Rightarrow B$ právě, když $T \models A \Rightarrow B$.

Definice 1. *Teorie T se nazývá báze $\mathcal{D}$, pokud pro každou $A \Rightarrow B$ platí $\mathcal{D} \models A \Rightarrow B$ p.k. $T \models A \Rightarrow B$.*

Poznámka. *Bází $\mathcal{D}$ je obecně hodně. Např. pokud T je báze $\mathcal{D}$ a navíc $T \models A \Rightarrow B$ pro nějakou $A \Rightarrow B \notin T$, pak $T \cup \{A \Rightarrow B\}$ je opět báze.*

Z definice báze je zřejmé, že budeme-li mít dvě báze, budou mít stejné sémantické důsledky, je proto žádoucí si takový jev pojmenovat.

Definice 2. *Teorie T_1 a T_2 jsou sémanticky ekvivalentní, značeno $T_1 \equiv T_2$, jestliže pro libovolnou $A \Rightarrow B$ platí $T_1 \models A \Rightarrow B$ právě, když $T_2 \models A \Rightarrow B$.*

Sémanticky ekvivalentní teorie, pak mají úzký vztah k pojmu model teorie.

Věta 1 (o charakterizaci sémantické ekvivalence)**.** *Následující tvrzení jsou ekvivalentní:*

1. $T_1 \equiv T_2$,
2. $\mathrm{Mod}(T_1) = \mathrm{Mod}(T_2)$,
3. $\mathrm{Mod}_C(T_1) = \mathrm{Mod}_C(T_2)$,
4. *Pro libovolnou $A \subseteq R$ máme $[A]_{T_1} = [A]_{T_2}$.*

Důkaz. $1 \Rightarrow 2$: Pro libovolnou $A \Rightarrow B$ máme $\mathrm{Mod}(T_1) \models A \Rightarrow B$ p.k. $T_1 \models A \Rightarrow B$ p.k. $T_2 \models A \Rightarrow B$ p.k. $\mathrm{Mod}(T_1) \models A \Rightarrow B$.

$2 \Rightarrow 3$: Speciální případ.

$3 \Rightarrow 4$: Stejné uzávěrové systémy mají stejné uzávěrové operátory.

$4 \Rightarrow 1$: Pro libovolnou $A \Rightarrow B$ máme $T_1 \models A \Rightarrow B$ p.k. $B \subseteq [A]_{T_1} = [A]_{T_2}$ p.k. $T_2 \models A \Rightarrow B$. □

Důsledek. *Pokud jsou T_1 a T_2 báze $\mathcal{D}$, pak $T_1 \equiv T_2$.*

Pro snadnější charakterizaci pravdivosti v relaci si zavedeme operátor, který bude fungovat podobně jako sémantický uzávěr u teorie. Nejdříve však definujeme relaci na n-ticích.

Definice 3. *Pro $\mathcal{D} \subseteq \prod_{y \in R} D_y$ a $M \subseteq R$ definujeme $E_{\mathcal{D}} : 2^R \to 2^{\mathcal{D} \times \mathcal{D}}$ předpisem*

$$E_{\mathcal{D}}(M) = \{\langle t, t' \rangle \in \mathcal{D} \times \mathcal{D} \mid t(M) = t'(M)\}.$$

Poznámka.
- *Z definice $E_{\mathcal{D}}$ je hned zřejmé, že relace $E_{\mathcal{D}}(M)$ je ekvivalencí na $\mathcal{D}$ což také znamená, že můžeme udělat rozklad.*

- *Význam vztahu $E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(M')$ je, že všechny dvojce n-tic, které se rovnají na M se také rovnají na M'.*
- *$E_{\mathcal{D}}$ je zřejmě antinonní, protože pokud $M_1 \subseteq M_2$, všechny n-tice, které se rovnají na M_2 se tím spíš musí rovnat na M_1, tedy $E_{\mathcal{D}}(M_2) \subseteq E_{\mathcal{D}}(M_1)$.*

Definice 4. *Pro $\mathcal{D} \subseteq \prod_{y \in R} D_y$ a $M \subseteq R$ definujeme $C_{\mathcal{D}} : 2^R \to 2^R$ předpisem*

$$C_{\mathcal{D}}(M) = \{y \in R \mid E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(\{y\})\}.$$

Poznámka. *$C_{\mathcal{D}}(M)$ je vlastně množina atributů, na kterých jsou si rovny všechny dvojice n-tic z $\mathcal{D}$, které jsou si rovny na M. Důsledkem pak je, že $E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(C_{\mathcal{D}}(M))$. Důkaz je ponechán čtenáři.*

Věta 2. *$C_{\mathcal{D}}$ je uzávěrový operátor na R.*

Důkaz.

- *(extenzivita)*: Pokud $y \in M$, pak $E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(\{y\})$, protože pokud jsou si t, t' rovny na všech atributech z M, tím spíš jsou si rovny na $y \in M$. Odtud dle definice $C_{\mathcal{D}}$ dostáváme $y \in C_{\mathcal{D}}(M)$.
- *(monotonie)*: Předpokládejme $M_1 \subseteq M_2$ a vezmeme $y \in C_{\mathcal{D}}(M_1)$. Poslední znamená, že $E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(\{y\})$. Z antitonie $E_{\mathcal{D}}$ dostáváme $E_{\mathcal{D}}(M_2) \subseteq E_{\mathcal{D}}(M_1) \subseteq E_{\mathcal{D}}(\{y\})$. Z definice $C_{\mathcal{D}}$ je $y \in C_{\mathcal{D}}(M_2)$.
- *(idempotence)*: $C_{\mathcal{D}}(M) \subseteq C_{\mathcal{D}}(C_{\mathcal{D}}(M))$ platí z extenzivity. Pro obrácenou inkluzi máme následující posloupnost argumentů:

$$E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(C_{\mathcal{D}}(M))$$
$$\{y \in R \mid E_{\mathcal{D}}(C_{\mathcal{D}}(M)) \subseteq E_{\mathcal{D}}(\{y\})\} \subseteq \{y \in R \mid E_{\mathcal{D}}(M) \subseteq E_{\mathcal{D}}(\{y\})\}$$
$$C_{\mathcal{D}}(C_{\mathcal{D}}(M)) \subseteq C_{\mathcal{D}}(M)$$

□

Věta 3 (o charakterizaci pravdivosti)**.** *Následující jsou ekvivalentní:*

1. $\mathcal{D} \models A \Rightarrow B$
2. $E_{\mathcal{D}}(A) \subseteq E_{\mathcal{D}}(B)$
3. $B \subseteq C_{\mathcal{D}}(A)$

Důkaz. 1 ⇒ 2: Z definice $\mathcal{D} \models A \Rightarrow B$, pokud $t(A) = t'(A)$, pak $t(B) = t'(B)$, tzn. pokud $\langle t, t' \rangle \in E_{\mathcal{D}}(A)$, pak $\langle t, t' \rangle \in E_{\mathcal{D}}(B)$, tj. $E_{\mathcal{D}}(A) \subseteq E_{\mathcal{D}}(B)$.

2 ⇒ 3: Předpokládejme $E_{\mathcal{D}}(A) \subseteq E_{\mathcal{D}}(B)$. Pro libovolný $y \in B$ pak z antitonie platí $E_{\mathcal{D}}(A) \subseteq E_{\mathcal{D}}(B) \subseteq E_{\mathcal{D}}(\{y\})$. To podle definice $C_{\mathcal{D}}$ znamená, že $y \in C_{\mathcal{D}}(A)$, tj. $B \subseteq C_{\mathcal{D}}(A)$.

3 ⇒ 1: Předpokládejme $B \subseteq C_{\mathcal{D}}(A)$. Dále mějme $t, t' \in \mathcal{D}$ takové, že $t(A) = t'(A)$ a vezmeme libovolné $y \in B$. Pak nutně $\langle t, t' \rangle \in E_{\mathcal{D}}(A)$ a navíc $E_{\mathcal{D}}(A) \subseteq E_{\mathcal{D}}(\{y\})$. Důsledkem je, že $t(y) = t'(y)$, tedy $t(B) = t'(B)$. □

Věta 4 (o charakterizaci báze)**.** *T je báze $\mathcal{D}$ právě, když pro libovolné $A \subseteq R$ máme $C_{\mathcal{D}}(A) = [A]_T$.*

Poznámka. *Ekvivalentně $C_{\mathcal{D}}(M) = [M]_T = M_T^{\infty} = M_T^{+}$.*

Důkaz. "$\Rightarrow$": Nechť T je báze $\mathcal{D}$. Pak $[M]_T \subseteq [M]_T$ p.k. $T \models M \Rightarrow [M]_T$ p.k. $\mathcal{D} \models M \Rightarrow [M]_T$ p.k. $[M]_T \subseteq C_{\mathcal{D}}(M)$. Obráceně máme $C_{\mathcal{D}}(M) \subseteq C_{\mathcal{D}}(M)$ p.k. $\mathcal{D} \models M \Rightarrow C_{\mathcal{D}}(M)$ p.k. $T \models M \Rightarrow C_{\mathcal{D}}(M)$ p.k. $C_{\mathcal{D}}(M) \subseteq [M]_T$. Dohromady tedy $C_{\mathcal{D}}(M) = [M]_T$.

"$\Leftarrow$": Pokud $C_{\mathcal{D}}$ má stejné pevné body jako $[\dots]_T$, pak $\mathcal{D} \models A \Rightarrow B$ p.k. $B \subseteq C_{\mathcal{D}}(A) = [A]_T$ p.k. $T \models A \Rightarrow B$. □

Následující věta ukazuje, že pro libovolnou relaci existuje minimálně jedna báze.

Věta 5 (o existenci báze)**.** *$T = \{A \Rightarrow C_{\mathcal{D}}(A) \mid A \subseteq R\}$ je báze $\mathcal{D}$.*

Důkaz. Dle předchozí věty stačí ověřit, že $C_{\mathcal{D}}(M) = [M]_T$ pro libovolnou M, tzn. ověřit, že $M = C_{\mathcal{D}}(M)$ právě když $M \in \mathcal{M}_T$.

"$\Rightarrow$": Předpokládejme $M = C_{\mathcal{D}}(M)$ a vezmeme libovolnou $A \Rightarrow C_{\mathcal{D}}(A) \in T$ tak, že $A \subseteq M$. Z monotonie operátoru $C_{\mathcal{D}}$ dostaneme $C_{\mathcal{D}}(A) \subseteq C_{\mathcal{D}}(M) = M$. Dohromady tedy $\mathcal{D}_M \models A \Rightarrow C_{\mathcal{D}}(A)$ p.k. $\mathcal{D}_M \in \mathrm{Mod}_C(T)$ p.k. $M \in \mathcal{M}_T$.

"$\Leftarrow$": Předpokládejme, že $M \in \mathcal{M}_T$. To jest $\mathcal{D}_M \in \mathrm{Mod}_C(T)$. Speciálně pro $M \Rightarrow C_{\mathcal{D}}(M) \in T$ máme $\mathcal{D}_M \models M \Rightarrow C_{\mathcal{D}}(M)$. Odtud $C_{\mathcal{D}}(M) \subseteq M$ a přidáme-li extenzivitu $C_{\mathcal{D}}$ dostaneme $C_{\mathcal{D}}(M) = M$. □

Když už víme, že báze vždy existuje, přesuneme pozornost na její velikost vzhledem k počtu FZ. Z předchozího textu vyplývá, že se budeme snažit najít bázi ekvivalentní, ale co nejmenší.