

Logiikan uusimpia suuntauksia

3. Selitettävyyden perusteita

Aloitetaan selitettävyyden käsitteen käsittely tutustumalla lähteen [1] lukuun 3.1. Seuraavilla slideilla on tähän liittyviä kommentteja ja huomioita.

<https://trepo.tuni.fi/bitstream/handle/10024/150660/Lehtimaki.pdf?sequence=2&isAllowed=y>

[1] Ilpo Lehtimäki. Ilpo Lehtimäki. *Selitettävyydestä propositiologiikassa*. Tampereen yliopisto, 2023.

Luvun 3.1 kommentointia

Huom., määritelmässä 3.1.4 on painovirhe. Määritelmän korjattu versio:

Olkoon $(\mathcal{M}, \mathcal{F}, v, m)$ logiikka. Määritellään ekvivalenssirelaatio $\equiv \subseteq \mathcal{F} \times \mathcal{F}$ siten, että $(\varphi, \varphi') \in \equiv$ jos ja vain jos $v(M, \varphi) = v(M, \varphi')$ kaikille malleille $M \in \mathcal{M}$.

Luvun 3.1 kommentointia

Esimerkki yleisestä selitysongelmasta. Tarkastellaan logiikkaa $(\mathcal{M}, \mathcal{F}, v, m)$, joka määrittellään seuraavasti.

1. Olkoon $U = \{x, y, z, \dots\}$ numeroituvasti ääretön *muuttujien* joukko. Olkoon joukko $\mathcal{M}$ kokonaislukuarvoisten **sijoitusfunktioiden** joukko, eli siis funktioiden

$$f : U \rightarrow \mathbb{Z}$$

joukko.

2. $\mathcal{F}$ on usean muuttujan kokonaislukukertoimisten polynomien joukko. Esimerkiksi siis vaikkapa

$$x^2y^3 + xy^2 - x - 6$$

kuuluu joukkon $\mathcal{F}$.

Luvun 3.1 kommentointia

3. v on funktio $v : \mathcal{M} \times \mathcal{F} \rightarrow \mathbb{Z}$ siten, että $v(f, p)$ antaa polynomin p arvon sijoituksella f . Esimerkiksi vaikkapa

$$v(x^2y, \{(x, 2), (y, 3) \dots\}) = 2^2 \cdot 3 = 12$$

4. $m : \mathcal{F} \rightarrow \mathbb{N}$ antaa syötteellä $p \in \mathcal{M}$ polynomin p pituuden, eli merkkien lukumäärän. (Emme määrittele pituutta tässä tarkemmin, sillä se ei ole oleellista esimerkin kannalta.)

Luvun 3.1 kommentointia

Tämän logiikan lauseet siis ovat polynomeja ja semantiikkafunktio v antaa sijoitusfunktiolle f ja polynomille p kyseisen polynomin arvon joukossa $\mathbb{Z}$.

Tarkastellaan tämän logiikan yleistä selitysongelmaa.

- ▶ Syötteellä $2x - x$ saadaan tulosteeksi x .
- ▶ Syötteellä

$$x^2y - 3x^2 - y^2x - 2x^2y + x^2 + yx^2 + y^2x + 3x^2$$

saadaan tulosteeksi x^2

- ▶ Syötteellä $x + x + x + x + x$ tulosteena on $5x$, missä siis $5x$ on kahden merkin pituinen polynomi.

Luvun 3.1 kommentointia

Esimerkkejä erityisestä selitysongelmasta

Syöte:

$$(\{(p_1, 0), (p_2, 1), (p_3, 1)\}, p_1 \vee p_2 \vee (p_3 \wedge \neg p_3), 1)$$

eli "miksi $p_1 \vee p_2 \vee (p_3 \wedge \neg p_3)$ on totta kun $p_1 = 0$ ja muut propositiot totta?"

Tuloste:

p_2

eli "siksi, että p_2 on totta."

Luvun 3.1 kommentointia

Esimerkkejä erityisestä selitysongelmasta

Syöte:

$(\{(p_1, 1), (p_2, 0), (p_3, 0), (p_4, 0)\}, \neg(\neg p_1 \vee (p_2 \wedge p_1)) \vee (p_3 \vee \neg p_1 \vee p_4), 1)$

eli "miksi $\neg(\neg p_1 \vee (p_2 \wedge p_1)) \vee (p_3 \vee \neg p_1 \vee p_4)$ on totta kun $p_1 = 1$ ja muut propositiot 0?"

Tuloste:

$p_1 \wedge \neg p_2$

eli "siksi, että $p_1 = 1$ ja $p_2 = 0$."

Luvun 3.1 kommentointia

Huom., lähteen [1] sivun 17 alusta alkavaan esimerkkiin lisättäköön seuraava huomautus. *Myös $p_i \vee p_j$ on selvästi kelpaamaton selityskaavaksi, sillä se on tosi kaikilla valuaatioilla, joilla p_i tai p_j on tosi. Tällaisia valuaatioita esiintyy joukossa $\mathcal{M} = \mathcal{V} \setminus \{(p_1, 1), \dots, (p_n, 1)\}$. Todettakoon vielä, että $\neg\neg p_i$ ei kelpaa selitykseksi samasta syystä, kuin p_i .*

Selitettävyyden mahdollistamiseksi voidaan totta kai hyödyntää myös muita logiikoita propositionaalisen logiikan lisäksi. Vaikka rajoituttaisiin tilanteisiin, joissa syötejoukko S koostuu bittijonoista pituudeltaan k , toimii myös ensimmäisen kertaluvun logiikka melko luonnollisesti. Ideana on käyttää hyväksi sanamalleja. Myös useita muita logiikoita on helppo hyödyntää. Yleisenä kysymys on, millaiset logiikat ovat hyödyllisiä nimenomaan tekoälysovellusten yhteydessä.

Lähteet

- [1] Ilpo Lehtimäki. Ilpo Lehtimäki. *Selitettävyydestä propositiologiikassa*. Tampereen yliopisto, 2023.
- [2] Reijo Jaakkola, Tomi Janhunen, Antti Kuusisto, Masood F. Rankooh, and Miikka Vilander. *Explainability via short formulas: the case of propositional logic with implementation*. In RCRA 2022, CEUR Workshop Proceedings, pages 64–77. CEUR-WS.org, 2022.