

# **Rule-based Knowledge Graph Completion with Canonical Models**

**Zināšanu grafu pabeigšana ar noteikumiem balstītiem  
kanoniskiem modeļiem**

Ronalds Rundāns 2024

# Izmantotā literatūra

## Rule-based Knowledge Graph Completion with Canonical Models

Simon Ott\*

Bosch Center for Artificial  
Intelligence  
Renningen, Germany  
Simon.Ott@de.bosch.com

Patrick Betz

University Mannheim  
Mannheim, Germany  
patrick@informatik.uni-  
mannheim.de

Daria Stepanova

Bosch Center for Artificial  
Intelligence  
Renningen, Germany  
Daria.Stepanova@de.bosch.com

Mohamed H. Gad-Elrab

Bosch Center for Artificial  
Intelligence  
Renningen, Germany  
Mohamed.Gad-Elrab@de.bosch.com

Christian Meilicke

University Mannheim  
Mannheim, Germany  
christian@informatik.uni-  
mannheim.de

Heiner Stuckenschmidt

University Mannheim  
Mannheim, Germany  
heiner@informatik.uni-  
mannheim.de

# Zināšanu grafi

- **Struktūra, kas parāda zināšanas par kādu domēnu.**
- **Freebase**
- **DBPedia**
- **Google Knowledge Graph**
- **Microsoft Satori**

# Bieži nepilnīgi

- Miljardiem elementu un saišu
- Knowledge graph completion (KGC) or link prediction

# Risanāšanas metodes

- Deep Learning
- Ar noteikumiem - bāzēta pieeja – var izsekot katru minējumu līdz noteikumu ģenerēšanai un kādi fakti tos ietekmēja.

# Grafa piemērs

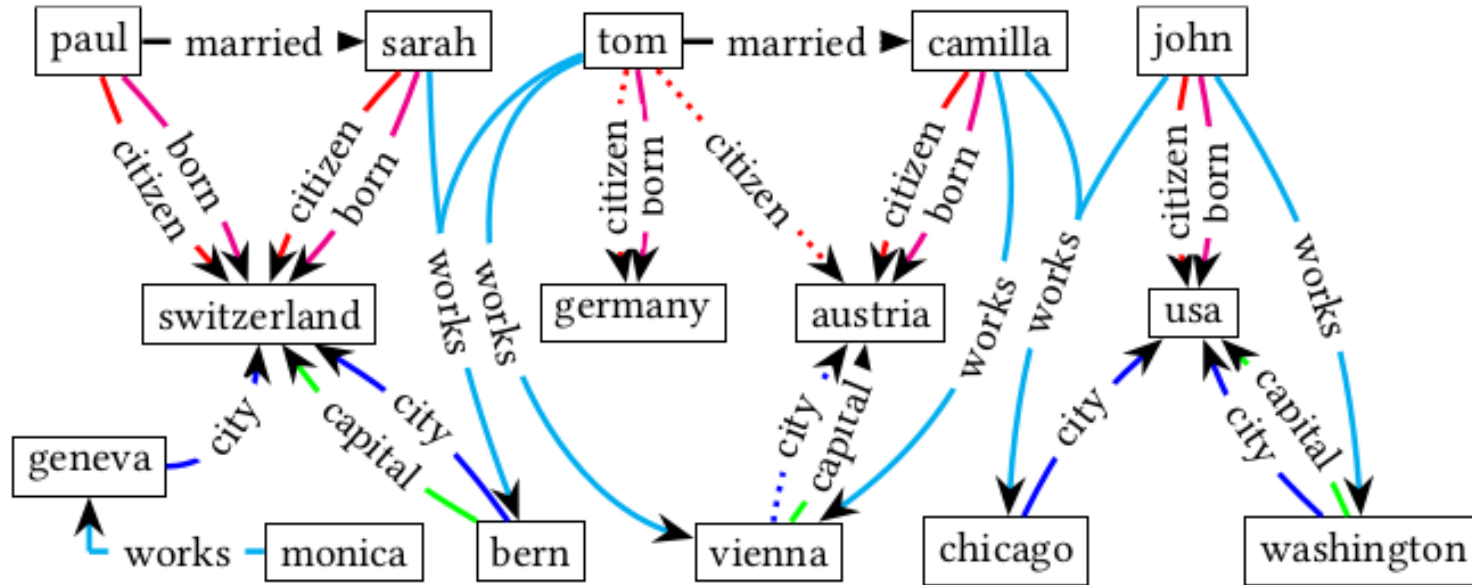


Figure 1: Example KG  $\mathcal{G}$  where solid edges represent known relations and dotted edges represent unknown relations.

# Problēma

$$\mathcal{G} \subset \{p(s, o) \mid p \in \mathcal{P}, s, o \in \mathcal{C}\}$$

- $\mathcal{C}$  ir konstanšu kopa (a.k.a. entities)
- $\mathcal{P}$  ir bināru predikātu kopa (a.k.a. relations)
- $s$  ir subjekts
- $o$  ir objekts
- $p(s, o)$  – (“triple”) fakts

# Kāda ir Toma pilsonība?

- **Vaicājums** `citizen(tom,?)`
- `citizen(tom,germany)` vai `citizen(tom, switzerland)` ?
- Šiem vaicājumiem piešķirs varbūtību skaitļus, lai tos sarindotu
- Automatizēta procedūra trūkstošo KG faktu konstatēšanai



# Pabeigšanas vaicājums

- Pabeigšanas vaicājumu parasti iegūst no testa “triple” kas ir viens no testa kopas “tripl - iem”.
- Katrs “triple” sastāv no diviem vaicājumiem, pa vienam katram virzienam.

# Testa kopa

- Ja vaicājums `citizen(tom,germany)` ir testa kopa, tad jāatbild uz sekojošiem vaicājumiem:  
`citizen(tom,?)` un `citizen(?,germany)` .
- Attiecīgajam modelim ir jāprognozē un jāvērtē fakti tādi, ka lielāka vērtība dažām prognozēm nozīmē, ka tā ir ticamāka nekā prognoze ar mazāku vērtību.

# Kas ir uz noteikumiem balstītas pieejas?

- Noteikumu bāzes pieejas izpilde ir būtiski atkarīga no noteikumu apkopošanas(agregācijas) problēmas risinājuma, kas attiecas uz punktu skaita aprēķināšanu - prognoze, ko ģenerē vairākas kārtulas.

# Kas ir uz noteikumiem balstītas pieejas?

- Uz noteikumiem balstītās KGC metodes skaidri atspoguļo pareizību, kas atrodama KG.
- Šos attēlojumus sauc par kārtulām, loģikas formulām vai klauzulām.
- Tālāk ir parādīti daži noteikumi, ko var iegūt no piemēra KG.

## Noteikumi no Example KG

$r_1 : \text{citizen}(X, Y) \leftarrow \text{born}(X, Y)$	0.80[4/5]
$r_2 : \text{citizen}(X, Y) \leftarrow \text{works}(X, A), \text{city}(A, Y)$	0.60[3/5]
$r_3 : \text{citizen}(X, Y) \leftarrow \text{works}(X, A), \text{capital}(A, Y)$	0.60[3/5]
$r_4 : \text{citizen}(X, Y) \leftarrow \text{married}(X, A), \text{citizen}(A, Y)$	0.50[1/2]

## Trīs noteikumu tipi

$$h(X, Y) \leftarrow b_1(X, A_1), b_2(A_1, A_2), \dots, b_n(A_n, Y)$$

$$h(X, c_h) \leftarrow b(X, c_d)$$

- Lielie burti – mainīgie.  $h(X, c_h) \leftarrow b(X, A)$
- mazie burti - konstantes,  $h(\dots)$  noteikuma galvas (head) atoms ar mērķa relāciju  $h$
- $b_1(\dots) \dots b_n(\dots)$  ir noteikuma ķermeņi (body) atomi.

## Noteikumi no Example KG

$r_1 : \text{citizen}(X, Y) \leftarrow \text{born}(X, Y)$	0.80 [4/5]
$r_2 : \text{citizen}(X, Y) \leftarrow \text{works}(X, A), \text{city}(A, Y)$	0.60 [3/5]
$r_3 : \text{citizen}(X, Y) \leftarrow \text{works}(X, A), \text{capital}(A, Y)$	0.60 [3/5]
$r_4 : \text{citizen}(X, Y) \leftarrow \text{married}(X, A), \text{citizen}(A, Y)$	0.50 [1/2]

## “Triples” no vaicājuma $\text{citizen}(\text{tom},?)$ ar noteikumiem no $r_1$ līdz $r_4$

$\mathcal{G} \cup r_1 \models \text{citizen}(\text{tom}, \text{germany})$

$\mathcal{G} \cup r_2 \models \text{citizen}(\text{tom}, \text{switzerland})$

$\mathcal{G} \cup r_3 \models \text{citizen}(\text{tom}, \text{switzerland}), \text{citizen}(\text{tom}, \text{austria})$

$\mathcal{G} \cup r_4 \models \text{citizen}(\text{tom}, \text{austria})$



## Ko piedāvā šajā rakstā?

- Ierosina uzraudzītu pieeju, lai uzzinātu pārsvērtu ticamības vērtību katram noteikumam, lai iegūtu optimālu skaidrojumu par apmācības kopumu, ņemot vērā konkrētu agregācijas funkciju.

# Max+ agregācija

- Sakārto prognozes leksikogrāfiskā secībā.
- Rezultāta pamatā ir noteikumu maksimums, kas deva prognozi.
- Ja šī vērtība ir identiska divām prognozēm, tad rekursīvi kārtojot prognozes tālāk, izmantojot nākamos augstākos noteikumus, kamēr nav izšķiramas visas prognozes līdz vairākiem top k skaitā vai līdz visu noteikumu piemērošanai.

# Noisy-or agregācija

- Modelis, lai raksturotu attiecības starp mainīgajiem Baijesas nosacītās varbūtības tabulā, un apraksta varbūtību, ka vismaz viens noteikums, kas paredzēja saiti, paredzēja to pareizi.
- $R_t \subseteq R$  ir prognozētie noteikumi no “triple”  $t$ , kam piešķir:

$$score(t) = 1 - \prod_{r \in R_t} (1 - conf(r))$$

# Noisy-or

- Mainīgie lielumi, kas parādās varbūtības tabulā, mūsu darba kontekstā atbilst apgūtajiem noteikumiem.
- Kārtulām saistītie svāri nav mācīti, tādēļ bieži sliktāki rezultāti nekā Max+.

# Noteikumu r uzticamība

- Noteikums  $r$  , tā uzticamība ( confidence) ir:  
 $\text{conf}(r) = TP / (TP + FP + Ic)$  , kur
- FP-False positives
- TP -true positives
- maza konstante  $Ic > 0$
- Relatīvs skaitlis – patiesās vērtības no visiem likuma minējumiem.

Melns = pareizi, sarkana = nepareizi

**Table 1: Rankings generated using Max+ and Noisy-or based on the rules  $r_1$  to  $r_4$ .**

	Max+	Noisy-or
1	germany (0.8)	switzerland (0.84)
2	switzerland (0.6,0.6)	germany (0.8)
3	austria (0.6,0.5)	austria (0.8)

# Rangu rezultāti

- Switzerland ir uzpūsta vērtība no r2 un r3
- Tā kā nav zināms, ka Vīne ir pilsēta Austrijā, r2 neprognozē pilsoni (Tom, Austria), kas neļauj Max+ ģenerēt pareizu vērtību.

# Metode

- Šajā rakstā autori mēģina atrisināt noteikumu ticamības apkopošanas problēmu, iegūstot modificētu ticamības punktu variantu ar Noisy-or modeli.



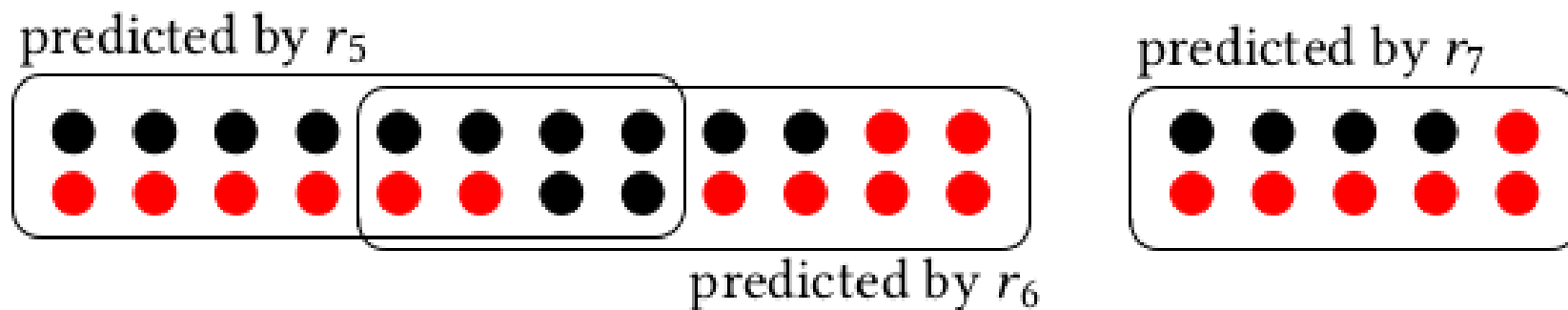
# Papildus noteikumi

$r_5 : \text{born}(X, Y) \leftarrow \text{married}(X, A), \text{born}(A, Y) \quad 0.625[10/16]$

$r_6 : \text{born}(X, \text{usa}) \leftarrow \text{gender}(X, \text{male}) \quad 0.5[8/16]$

$r_7 : \text{born}(X, \text{norway}) \leftarrow \text{works}(X, \text{oslo}) \quad 0.4[4/10]$

# Melns=eksistē, sarkans=neeksistē



**Figure 2: Predictions made by three rules  $r_5$ ,  $r_6$ , and  $r_7$  on example dataset  $D'$ . Black dots are predictions that exist in the KG, red dots are predictions that do not exist in the KG.**

**Table 2: Final values for a candidate to be true for different combinations of predicting rules and different aggregation strategies. NO and LR stand for Noisy-Or and Logistic Regression respectively.**

Rules			Data Optimal	w. Confidence		Learned Weights		
$r_5$	$r_6$	$r_7$		Max+	NO	LR	LR+	NO
			-	0.000	0.000	0.100	0.100	0.000
		×	0.400	0.400	0.400	0.400	0.400	0.400
	×		0.250	0.500	0.500	0.250	0.250	0.279
	×	×	-	0.500	0.700	0.667	0.667	0.567
×			0.500	0.625	0.625	0.500	0.500	0.557
×		×	-	0.625	0.775	0.857	0.857	0.734
×	×		0.750	0.625	0.813	0.750	0.750	0.680
×	×	×	-	0.625	0.888	0.947	0.947	0.808
BCE-Loss			0.272	0.292	0.289	0.272	0.272	0.273

# Noteikumu punktu interpretējamība

- **film genre (the\_love\_guru, ?)    =>    romantic\_comedy**

*genre(X, romantic\_comedy) ← genre(X, romance\_film)*                      1.09

*genre(X, romantic\_comedy) ← netflix\_genre(comedy, X)*                      1.08

*genre(X, romantic\_comedy) ← genre(X, comedy)*                      0.99

*genre(X, Y) ← distributor(A, X), nominated(A, B), genre(B, Y)*                      0.05

*genre(X, Y) ← composer(X, A), nominated(A, B), genre(B, Y)*                      0.04

# Secinājumi

- Lai uzlabotu noteikumu apkopošanas kvalitāti, apskatījām pieeju noteikumu uzticamības iegūšanu ar uzraudzību no datiem ar loģistisko regresiju.
- Ierosinātā metode parādīja labākus vai salīdzināmus rezultātus ar pašreizējiem neironu modeļiem, ja tos testēja pēc ievērojamiem etaloniem, lai gan izmantoja vienkāršākus modeļus.
- Svarīgi, ka ierosinātās prognožu vērtēšanas metodes ir elastīgas, jo tās var izmantot papildus ikvienai noteikumu apguves sistēmai.



**Jautājumi?**