Mining Symbolic Rules To Explain Lung Cancer Treatments

Disha Purohit
Prof. Dr. Maria-Esther Vidal

Example - Rule 1 (Adhering to clinical guidelines)

```
hasStage(IVB, ?x), hasOncologicalTreatment(Immunotherapy, ?x) → hasBio(PDL1, ?x)
```

Support - 28
Confidence - 0.2
PCA Confidence - 0.965517

- 28 instantiations of the body are also instantiations of the head, i.e.,
  - The support of this rule is 28.
  - For each of the 28 patients.
- 1 instantiation is identified as not following the clinical guidelines.

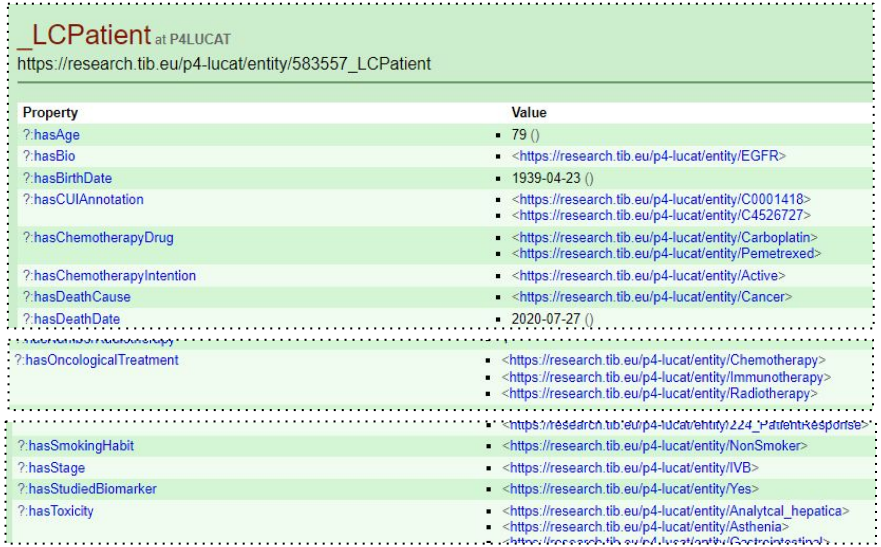
```
PREFIX p4-lucat: <http://research.tib.eu/p4-lucat/vocab/>
PREFIX p4-lucat_en: <http://research.tib.eu/p4-lucat/entity/>

SELECT DISTINCT ?s WHERE {
       ?s a p4-lucat:LCPatient .
       ?s p4-lucat:hasStage p4-lucat_en:IVB .
       ?s p4-lucat:hasOncologicalTreatment p4-lucat_en:Immunotherapy.
       ?s p4-lucat:hasBio ?bio .
       FILTER (?bio != p4-lucat_en:PDL1)
       FILTER (!EXISTS {?s p4-lucat:hasBio p4-lucat_en:PDL1})
       }
```

The following SPARQL query provides us the heuristic based negative edges of `hasBio(PDL1, ?x)`

For example, the patients
http://research.tib.eu/p4-lucat/entity/583557_LCPatient

Example - Rule 1 (Adhering to clinical guidelines)



_LCPatient at P4LUCAT
https://research.tib.eu/p4-lucat/entity/583557_LCPatient

Property	Value
?:hasAge	79 ()
?:hasBio	<https://research.tib.eu/p4-lucat/entity/EGFR>
?:hasBirthDate	1939-04-23 ()
?:hasCUIAnnotation	<https://research.tib.eu/p4-lucat/entity/C0001418> <https://research.tib.eu/p4-lucat/entity/C4526727>
?:hasChemotherapyDrug	<https://research.tib.eu/p4-lucat/entity/Carboplatin> <https://research.tib.eu/p4-lucat/entity/Pemetrexed>
?:hasChemotherapyIntention	<https://research.tib.eu/p4-lucat/entity/Active>
?:hasDeathCause	<https://research.tib.eu/p4-lucat/entity/Cancer>
?:hasDeathDate	2020-07-27 ()
?:hasOncologicalTreatment	<https://research.tib.eu/p4-lucat/entity/Chemotherapy> <https://research.tib.eu/p4-lucat/entity/Immunotherapy> <https://research.tib.eu/p4-lucat/entity/Radiotherapy>
?:hasSmokingHabit	<https://research.tib.eu/p4-lucat/entity/NonSmoker>
?:hasStage	<https://research.tib.eu/p4-lucat/entity/IVB>
?:hasStudiedBiomarker	<https://research.tib.eu/p4-lucat/entity/Yes>
?:hasToxicity	<https://research.tib.eu/p4-lucat/entity/Analytcal_hepatica> <https://research.tib.eu/p4-lucat/entity/Asthenia>

Clinical Guidelines	
Body	**Head**
Stage I	
(SUR? : Yes)	[CT, SUR]
(SUR? : No)	[RT, CT]
Stage II	
(SUR? : Yes)	[CT, SUR]
(SUR? : No)	[RT, CT]
Stage III	
IIIB (RES? : No)	[conc(CT+RT)]
III (RES? : No)	[seq(CT, RT)]
III (RES? : POT)	[conc(CT+RT), SUR]
IIIA (RES? : POT)	[conc(CT+RT), SUR]
Stage IV	
IV (MUT?: No)	[conc(Platimum-CT+Agent-DRUGS)]
IV (MUT? : Yes)	TKI
IV (PD-1/PD-L1? : Yes)	IMM

Lung cancer patient with medical record **583557** is in stage IVB and received oncological treatment immunotherapy but does not contain the biomarker 'PDL1'. Therefore, this patient potentially violates the clinical guidelines. Therefore, for this patient the missing link could correspond to prediction of `hasBio(PDL1, ?x)`

Example - Rule 2 (Identifying inaccuracy in data)

`hasOncologicalTreatment(Neoadjuvant, ?x) → hasOncologicalTreatment(Chemotherapy, ?x)`

Support - 34
Confidence - 0.001126
PCA Confidence - 0.971429

- 34 instantiations of the body are also instantiations of the head, i.e.,
  - The support of this rule is 34.
  - For each of the 34 patients.
- 1 instantiation is identified as not receiving chemotherapy treatment.

```
PREFIX p4-lucat: <http://research.tib.eu/p4-lucat/vocab/>
PREFIX p4-lucat_en: <http://research.tib.eu/p4-lucat/entity/>

SELECT DISTINCT ?s WHERE {
       ?s a p4-lucat:LCPatient .
       ?s p4-lucat:hasOncologicalTreatment p4-lucat_en:Neoadjuvant.
       ?s p4-lucat:hasOncologicalTreatment ?x .
       FILTER (?x != p4-lucat_en:Chemotherapy) .
       FILTER (!EXISTS
       {?s p4-lucat:hasOncologicalTreatment
       p4-lucat_en:Chemotherapy})}
```

The following SPARQL query provides us the heuristic based negative edges of `hasOncologicalTreatment(Chemotherapy, ?x)`

For example, the patients
http://research.tib.eu/p4-lucat/entity/1255806_LCPatient

Example - Rule 2 (Identifying inaccuracy in data)

_LCPatient at P4LUCAT
https://research.tib.eu/p4-lucat/entity/1255806_LCPatient

Property	Value
?:hasAge	52 ()
?:hasBirthDate	1963-01-14 ()
?:hasCUIAnnotation	<https://research.tib.eu/p4-lucat/entity/C0007137> <https://research.tib.eu/p4-lucat/entity/C0278984>
?:hasChemotherapyDrug	<https://research.tib.eu/p4-lucat/entity/Carboplatin> <https://research.tib.eu/p4-lucat/entity/Paclitaxel> <https://research.tib.eu/p4-lucat/entity/Vinorelbine>
?:hasChemotherapyIntention	<https://research.tib.eu/p4-lucat/entity/Active>
?:hasDeathCause	<https://research.tib.eu/p4-lucat/entity/Cancer>
?:hasDeathDate	2015-09-09 ()
?:hasDiagnosisDate	2014-12-18 ()
?:hasDrug	<https://research.tib.eu/p4-lucat/entity/Gastric_protector>
?:hasFamilyCancer	<https://research.tib.eu/p4-lucat/entity/Yes>
?:hasFirstMetastasisDiagnosis	<https://research.tib.eu/p4-lucat/entity/Desconocido>
?:hasGender	<https://research.tib.eu/p4-lucat/entity/Female>
?:hasIDTNMDiagnosis	<https://research.tib.eu/p4-lucat/entity/578>
is ?:hasID_LCPatient of	<https://research.tib.eu/p4-lucat/entity/1255806_1> <https://research.tib.eu/p4-lucat/entity/1255806_10> <https://research.tib.eu/p4-lucat/entity/1255806_15> <https://research.tib.eu/p4-lucat/entity/1255806_Father_C0684249> <https://research.tib.eu/p4-lucat/entity/1255806_Half_Sister_C0684249>
?:hasLastConsultationDate	2015-09-09 ()
?:hasNoOncologicalTreatment	<https://research.tib.eu/p4-lucat/entity/NoAdjuvant> <https://research.tib.eu/p4-lucat/entity/NoChemotherapy> <https://research.tib.eu/p4-lucat/entity/NoProphylactic> <https://research.tib.eu/p4-lucat/entity/NoRadical> <https://research.tib.eu/p4-lucat/entity/NoSurgery>
?:hasNonOncologicalDrug	<https://research.tib.eu/p4-lucat/entity/DB00338>
?:hasNumberCigarette	20
?:hasNumberLCRelatives	2
?:hasNumberOtherCancerRelatives	0
?:hasNumberRadiotherapy	1
?:hasOncologicalTreatment	<https://research.tib.eu/p4-lucat/entity/Neoadjuvant> <https://research.tib.eu/p4-lucat/entity/Radiotherapy>
?:hasOncologicalTreatmentLine	<https://research.tib.eu/p4-lucat/entity/1255806_222> <https://research.tib.eu/p4-lucat/entity/1255806_223>
?:hasPatientStatus	<https://research.tib.eu/p4-lucat/entity/Dead>
?:hasPatientWeight	49.0
?:hasPerformanceObservation	<https://research.tib.eu/p4-lucat/entity/1255806_2014-09-04> <https://research.tib.eu/p4-lucat/entity/1255806_2015-08-10> <https://research.tib.eu/p4-lucat/entity/1255806_2015-08-14> <https://research.tib.eu/p4-lucat/entity/1255806_2015-08-21> <https://research.tib.eu/p4-lucat/entity/1255806_2015-08-26>
?:hasPreviousCancers	<https://research.tib.eu/p4-lucat/entity/No>
?:hasRace	<https://research.tib.eu/p4-lucat/entity/Caucasian>
?:hasReceivedChemotherapy	<https://research.tib.eu/p4-lucat/entity/No>
?:hasReceivedRadiotherapy	<https://research.tib.eu/p4-lucat/entity/Yes>
?:hasReceivedSurgery	<https://research.tib.eu/p4-lucat/entity/No>
?:hasRegisteredTumorSize	0.0
?:hasSmokingHabit	<https://research.tib.eu/p4-lucat/entity/CurrentSmoker>
?:hasStage	<https://research.tib.eu/p4-lucat/entity/IIIB>

Lung cancer patient with medical record id **1255806** has received list of chemotherapy drugs but the data does not contain information about that patient receiving oncological treatment '*Chemotherapy*' .

Example - Rule 3 (Potential missing relationships)

```
hasStage(IB, ?x) → hasOncologicalSurgery(Lobectomy, ?x)
```

Support - 80
Confidence - 0.178174
PCA Confidence - 0.97561

- 80 instantiations of the body are also instantiations of the head, i.e.,
  - The support of this rule is 80.
  - For each of the 80 patients are in stage IB and receive lobectomy as surgery.
- 2 instantiation are in stage IB but does not receive ‘Lobectomy’ as surgery and are considered to be potential missing links in the KG.

```
PREFIX p4-lucat: <http://research.tib.eu/p4-lucat/vocab/>
PREFIX p4-lucat_en: <http://research.tib.eu/p4-lucat/entity/>

SELECT ?X WHERE {
       ?X p4-lucat:hasStage p4-lucat_en:IB .
       ?X p4-lucat:hasOncologicalSurgery ?Y1 .
       FILTER (?Y1!= p4-lucat_en:Lobectomy) .
       FILTER (!EXISTS {
       ?X p4-lucat:hasOncologicalSurgery p4-lucat_en:Lobectomy})}}
```

The following SPARQL query provides us the heuristic based negative edges of `hasOncologicalSurgery(Lobectomy, ?x)`

For example, the patients
http://research.tib.eu/p4-lucat/entity/2515307_LCPatient ,
http://research.tib.eu/p4-lucat/entity/942497_LCPatient
Did not receive ‘Lobectomy’ surgery.

Acknowledgements

- This work has been supported by the project **TrustKG**-Transforming Data in Trustable Insights with grant P99/2020 and the EraMed project **P4-LUCAT**

Thank you