Faithful Rule Learning and Extraction with Applications

Xiaxia Wang

Department of Computer Science, University of Oxford

Advisors: Prof. Ian Horrocks, Prof. Bernardo Cuenca Grau, Dr. David Tena Cucala

29th May 2025

UNIVERSITY OF OXFORD

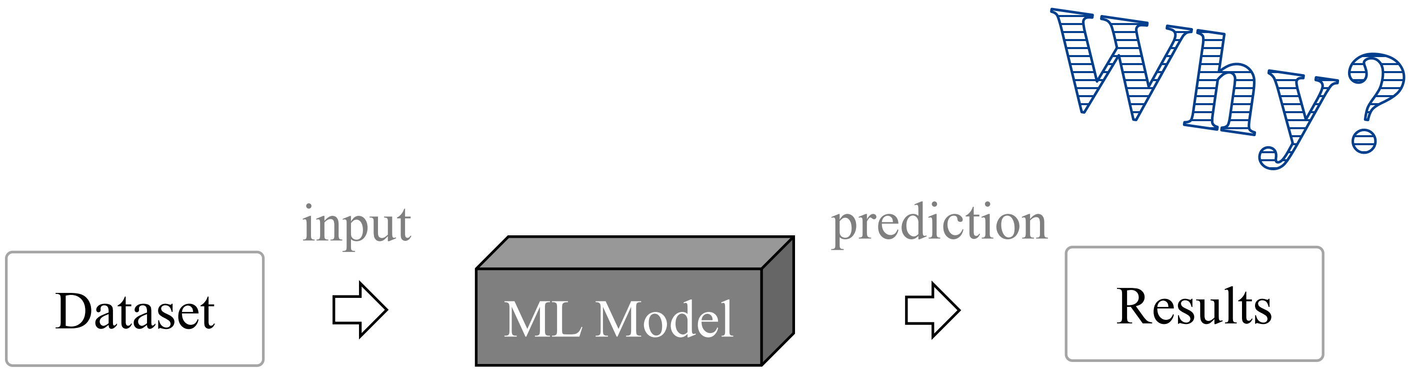
DEPARTMENT OF COMPUTER SCIENCE

Outline

1. Background
2. Motivation
3. Faithful Rule Learning for Knowledge Graph Completion
4. Faithful Rule Learning over Databases
5. Future Directions

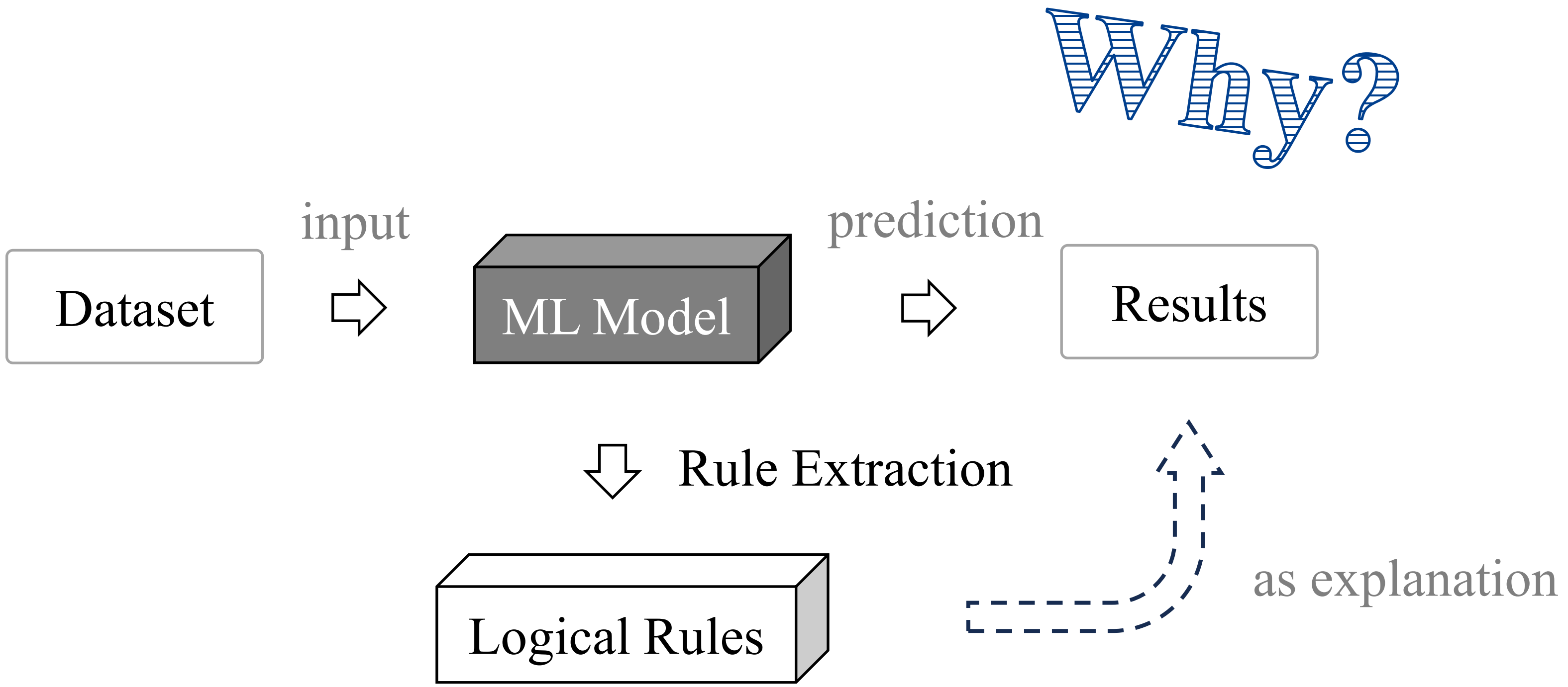
Background

- Machine Learning (ML) models are used in various domains, but they lack explainability



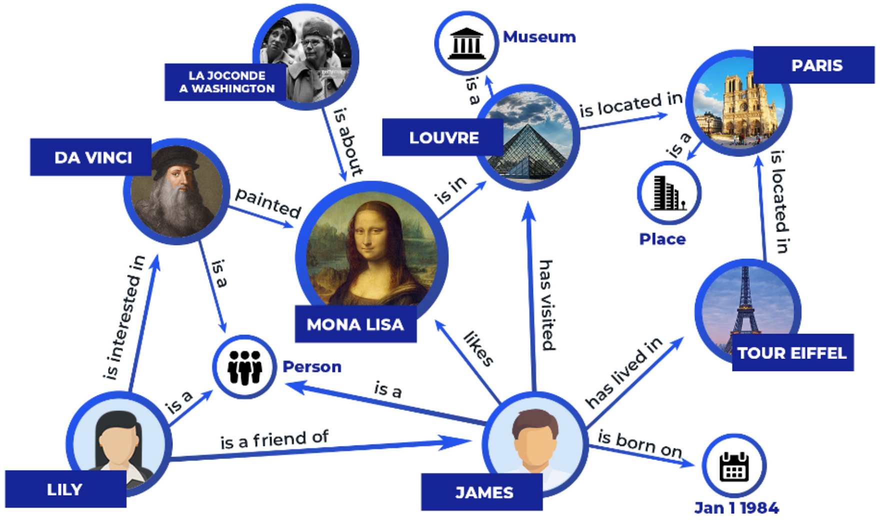
Background

- Machine Learning (ML) models are used in various domains, but they lack explainability
- Rule extraction improves the explainability of ML models



Background

- Machine Learning (ML) models are used in various domains, but they lack explainability
- Rule extraction improves the explainability of ML models
- We focus on ML models for graph data, e.g., knowledge graphs (KGs)



https://deeppavlov.ai/research/tpost/bn15u1y4v1-improving-knowledge-graph-completion-wit

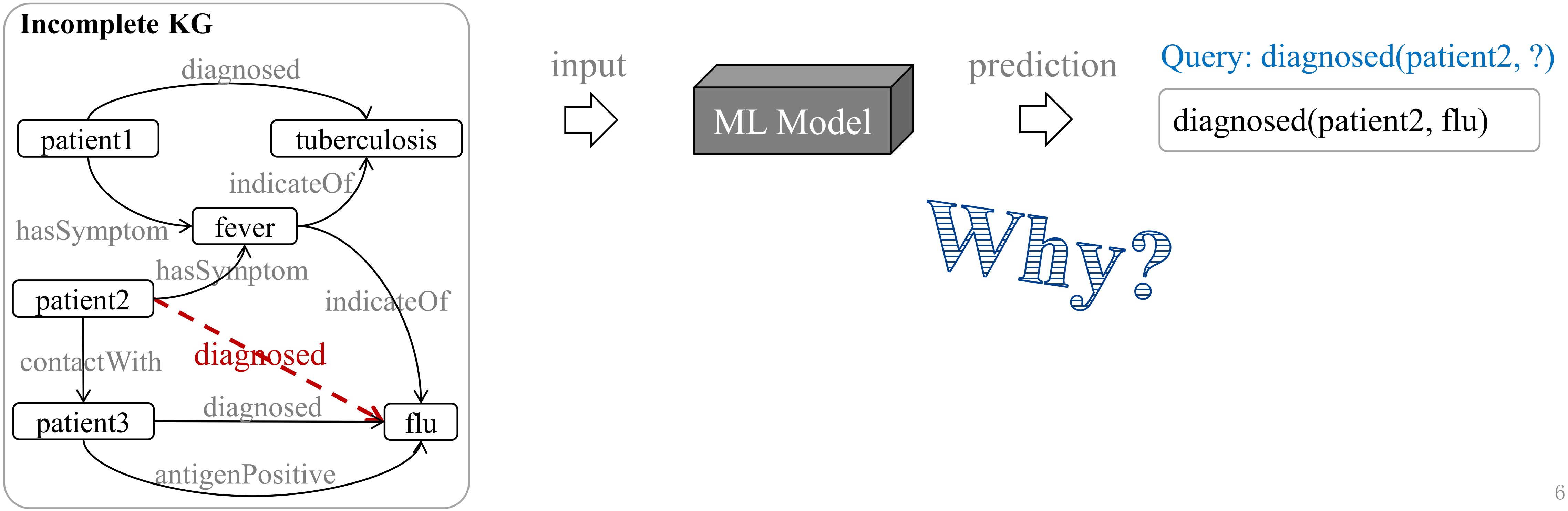
An example KG

Knowledge graph: a structured data representation

- each node: a real-world entity
- each directed edge: a relationship between entities, expressed as a fact *relation(subject, object)*, e.g., *painted(DA_VINCI, MONA_LISA)*
- a KG is a set of facts

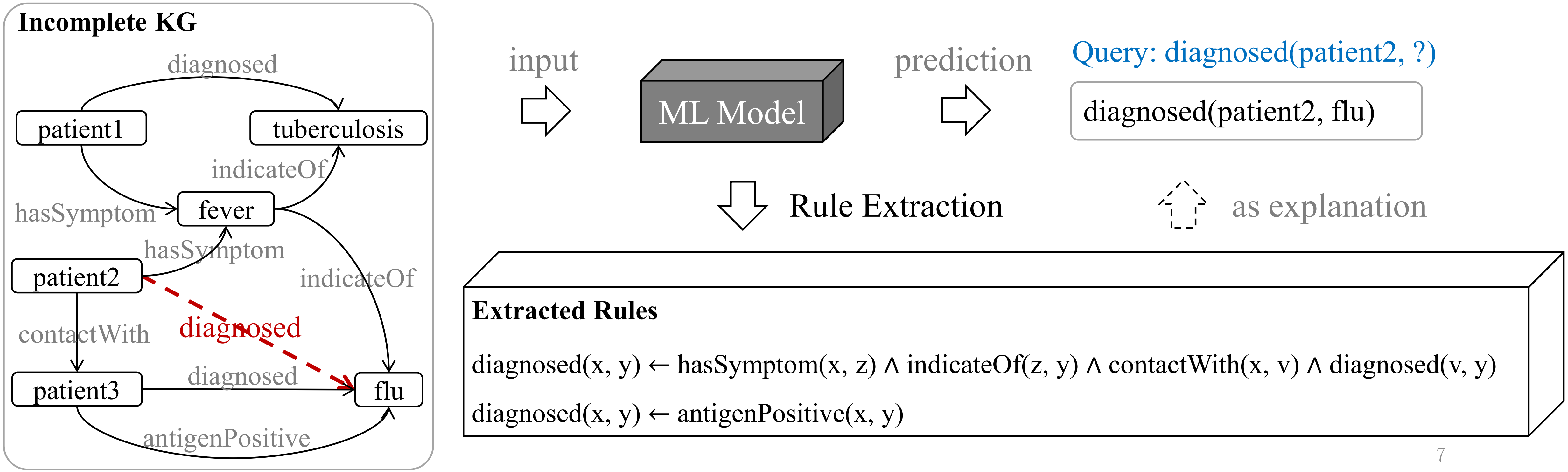
Background

- An example task: Knowledge Graph Completion (KGC)
  - real-world KGs are usually **incomplete**
  - KGC: to infer missing facts based on existing facts in the KG
  - given a query ***relation(subject, ?)***, to predict the object that matches ***?***



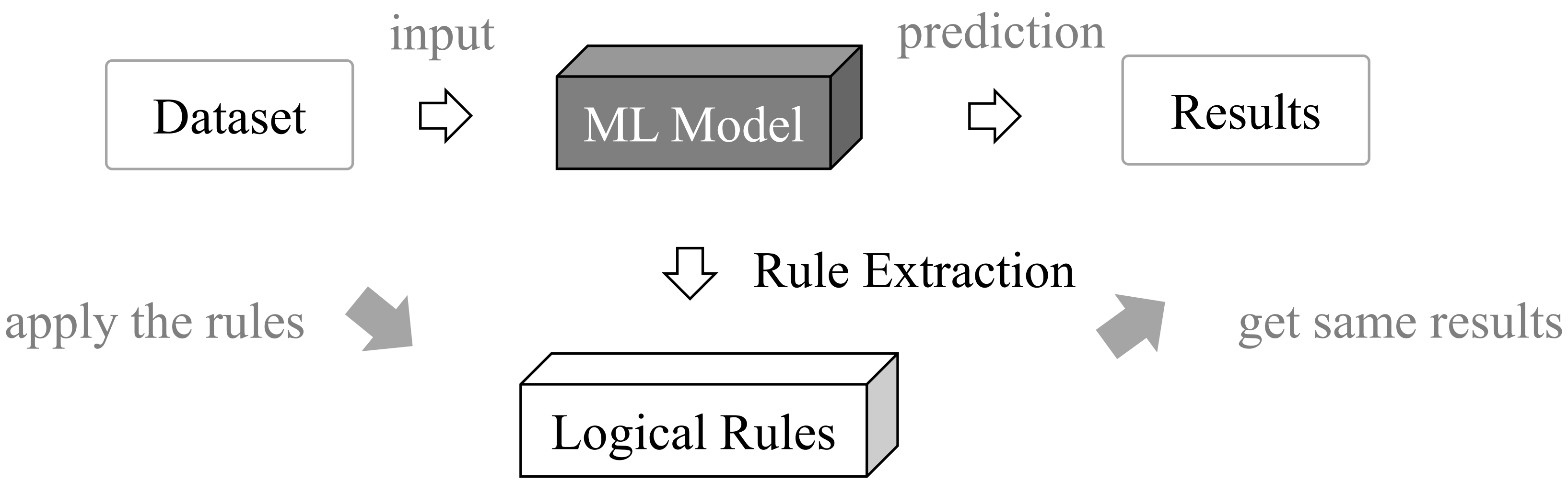
Background

- An example task: Knowledge Graph Completion (KGC)
- Rule extraction improves the explainability of ML models
  - by providing human-readable rules as explanation



Motivation

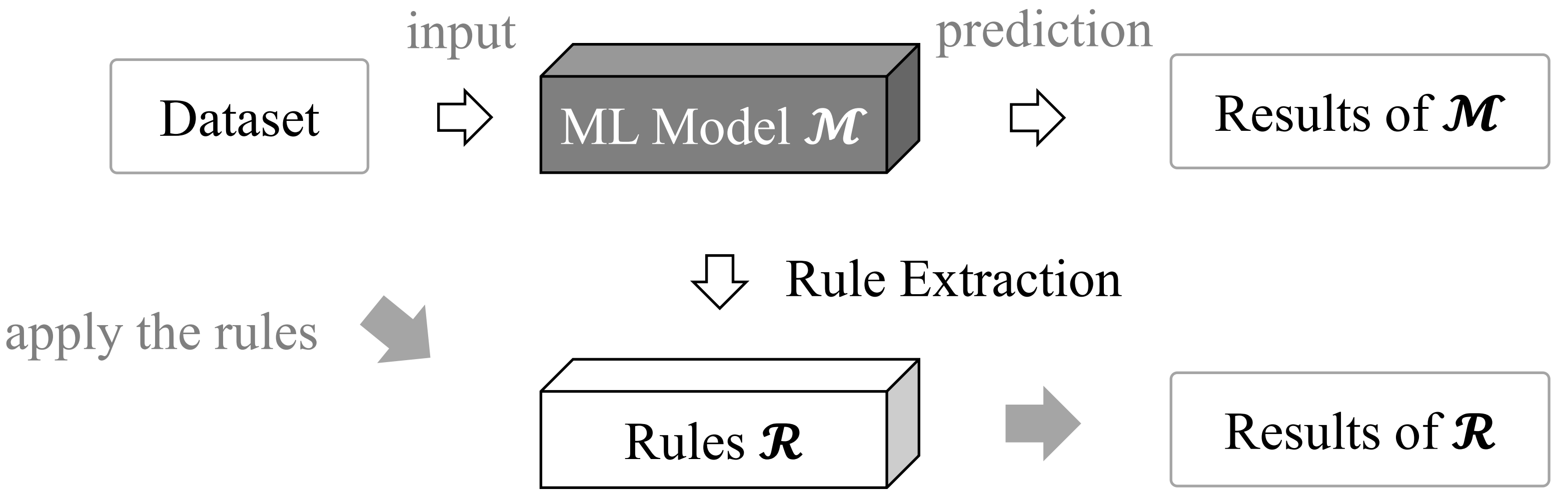
- Requirement: The extracted rules **must be “equivalent”** to the ML model
  - produce the same result on the same input
- Why is this important? — Otherwise, we cannot trust the rules to be explanation



*apply rules to a dataset: the fact (as a grounded rule head) is derived, if the rule body is grounded on the dataset

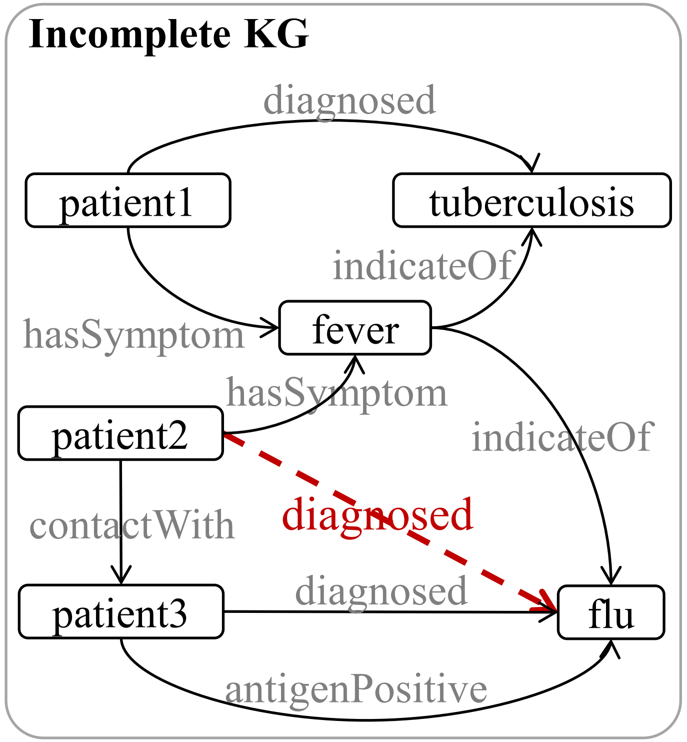
Motivation

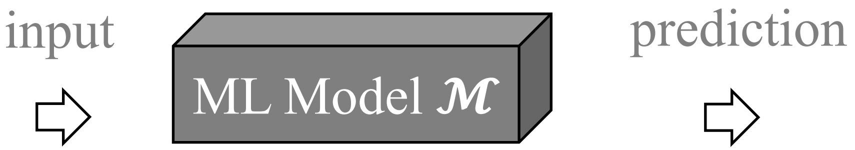
- Requirement: The extracted rules **must be "equivalent"** to the ML model
  - produce the same result on the same input
- Formally, let $\mathcal{M}$ and $\mathcal{R}$ denote the model and the set of extracted rules
  - $\mathcal{R}$ is sound for $\mathcal{M}$ if for any input dataset, {result of $\mathcal{R}$} $\subseteq$ {result of $\mathcal{M}$}
  - $\mathcal{R}$ is complete for $\mathcal{M}$ if for any input dataset, {result of $\mathcal{R}$} $\supseteq$ {result of $\mathcal{M}$}
  - $\mathcal{R}$ is faithful to $\mathcal{M}$ if both sound and complete



Motivation

- Requirement: The extracted rules **must be "equivalent"** to the ML model
  - produce the same result on the same input
- Issues with unsound or incomplete rules





Query: *diagnosed(patient2, ?)*

diagnosed(patient2, flu)

faithful rules $\mathcal{R}_1$

diagnosed(x, y) ← hasSymptom(x, z) ∧ indicateOf(z, y) ∧ contactWith(x, v) ∧ diagnosed(v, y)
diagnosed(x, y) ← antigenPositive(x, y)

Unsound rules $\mathcal{R}_2$

diagnosed(x, y) ← hasSymptom(x, z) ∧ indicateOf(z, y)

Wrong results

diagnosed(patient_2, flu)
diagnosed(patient_2, tuberculosis)

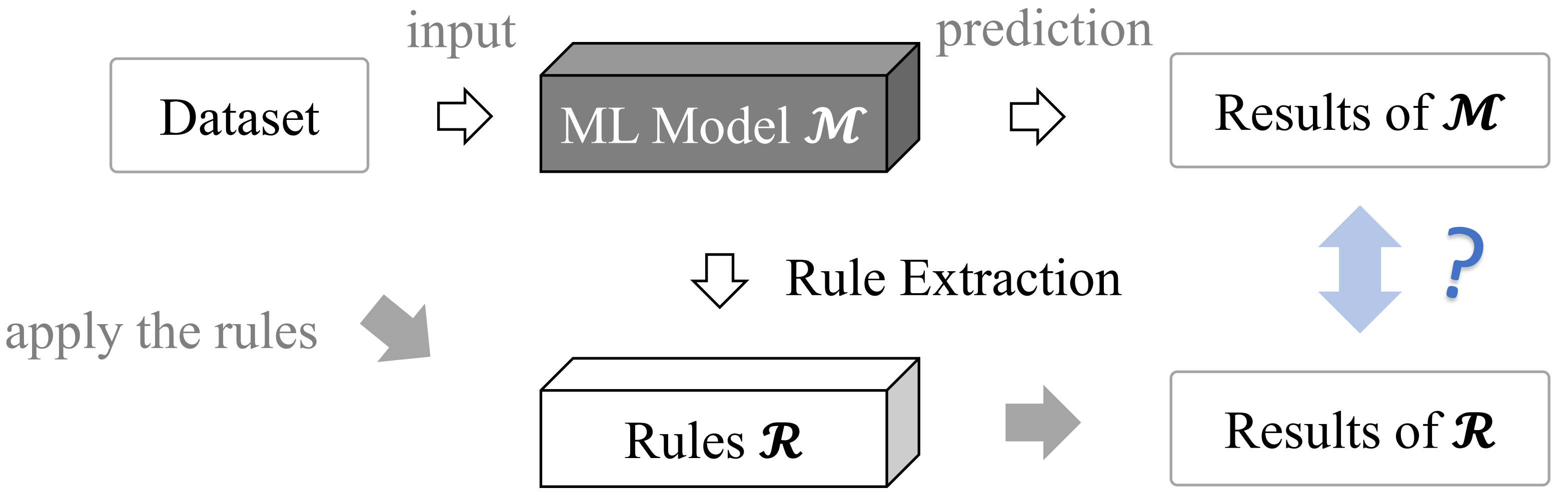
Incomplete rules $\mathcal{R}_3$

diagnosed(x, y) ← antigenPositive(x, y)

Missing results

Motivation

- Requirement: The extracted rules **must be "equivalent"** to the ML model
  - produce the same result on the same input
- Issues with unsound or incomplete rules
- Prior works[1—4] on rule extraction have **NO** formal guarantee of faithfulness

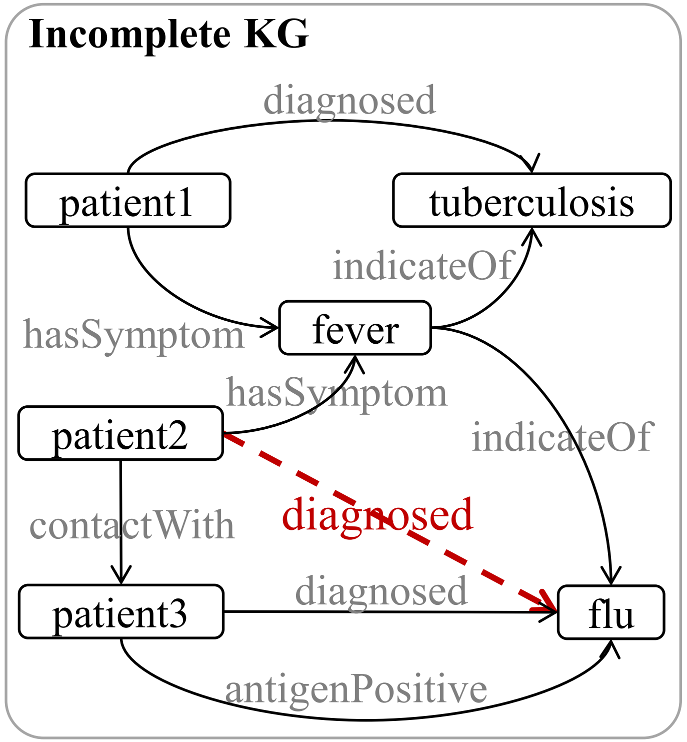


[1] Yang et al. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. NeurIPS 2017
[2] Sadeghian et al. DRUM: End-To-End Differentiable Rule Mining On Knowledge Graphs. NeurIPS 2019
[3] Xiong et al. TILP: Differentiable Learning of Temporal Logical Rules on Knowledge Graphs. ICLR 2023
[4] Han et al. Logical Entity Representation in Knowledge-Graphs for Differentiable Rule Learning. ICLR 2023

Faithful Rule Learning for Knowledge Graph Completion

- **Question 1**: Do existing approaches hold faithfulness?
- We consider DRUM[2], a representative rule learning approach for KGC
  - model: predict "missing" facts based on existing facts
  - rule extraction algorithm: obtain a set of chain-like rules

    e.g., *diagnosed(x, y)* ← *contactWith(x, z)* ∧ *diagnosed(z, y)*

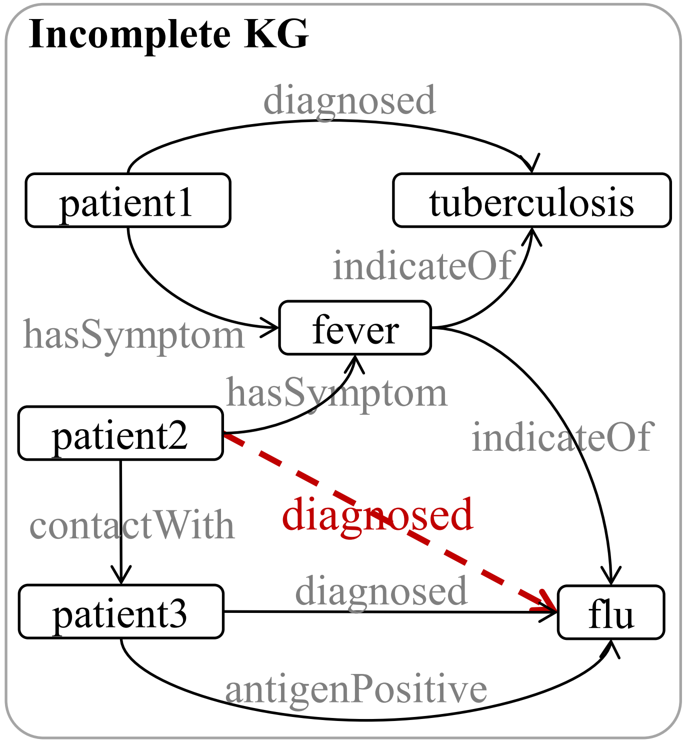


Faithful Rule Learning for Knowledge Graph Completion

UNIVERSITY OF OXFORD

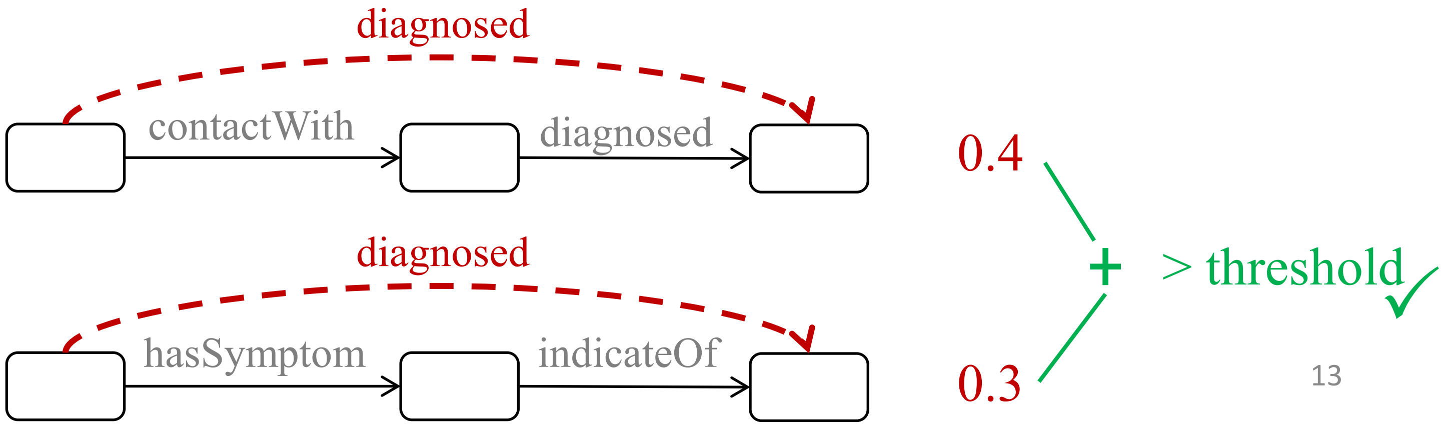
- **Question 1**: Do existing approaches hold faithfulness?
- We consider DRUM[2], a representative rule learning approach for KGC
  - model: predict “missing” facts based on existing facts
  - rule extraction algorithm: obtain a set of chain-like rules

    e.g., *diagnosed(x, y) ← contactWith(x, z) ∧ diagnosed(z, y)*



How does the model work?

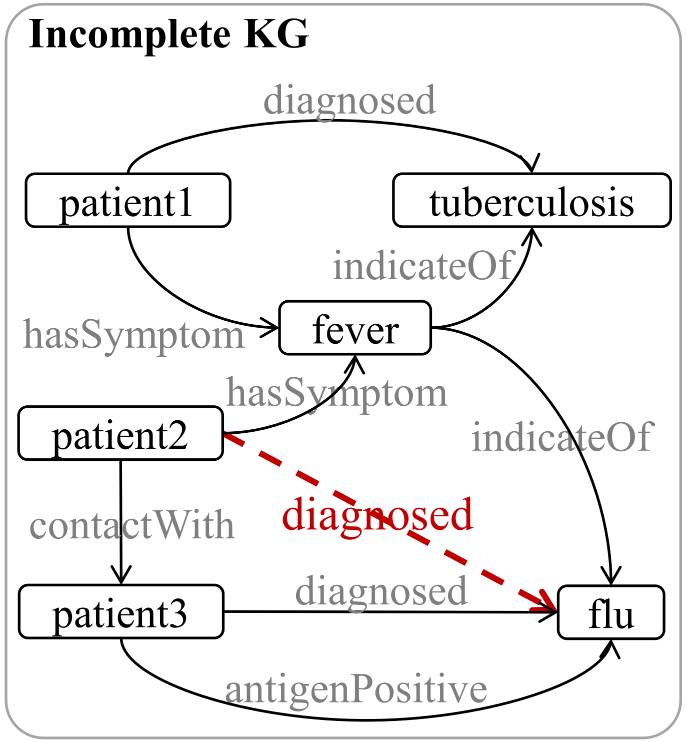
- for each relation, e.g., *diagnosed*
- it learns a weight for each “chain pattern”
- given a query *diagnosed(s, ?)*, it **sums up** all paths from *s* to the target node



- **Question 1**: Do existing approaches hold faithfulness?
- We consider DRUM[2], a representative rule learning approach for KGC
  - rule extraction algorithm: obtain a set of chain-like rules

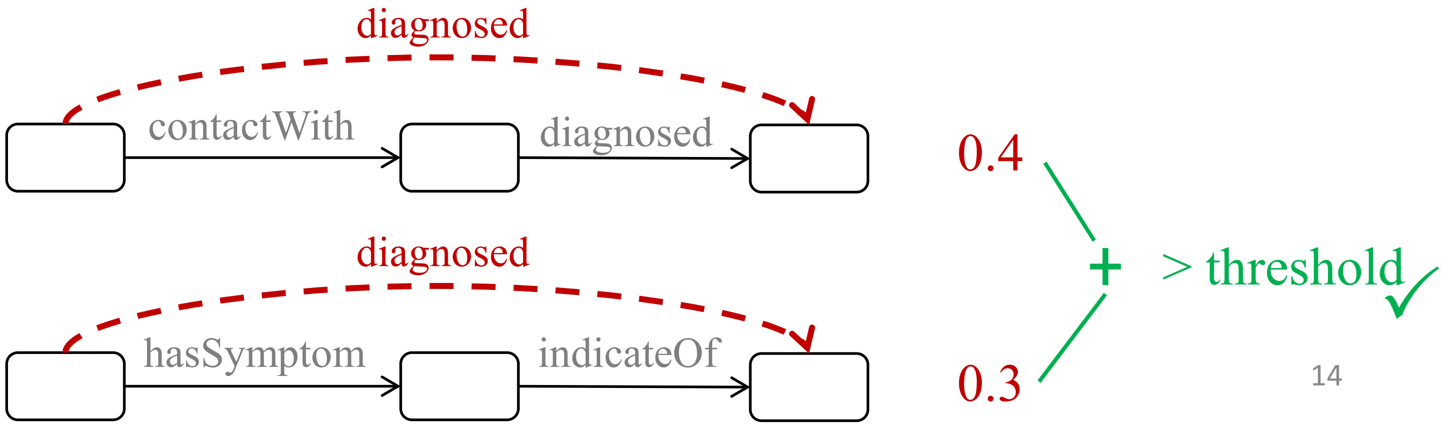
    e.g., *diagnosed(x, y)* ← *contactWith(x, z)* ∧ *diagnosed(z, y)*
  - **a single chain-like rule body cannot capture the "sum up" behavior**

NO

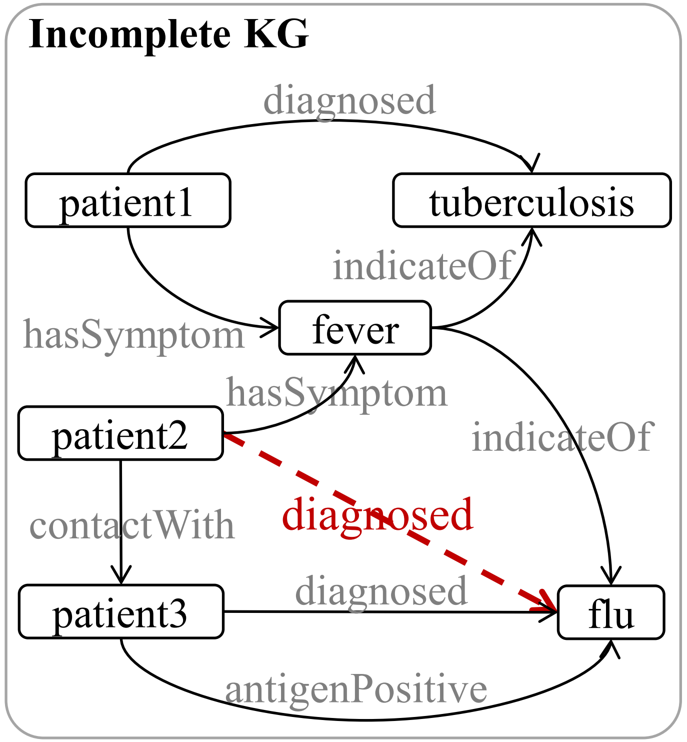


How does the model work?

- for each relation, e.g., *diagnosed*
- it learns a weight for each "chain pattern"
- given a query *diagnosed(s, ?)*, it **sums up** all paths from *s* to the target node

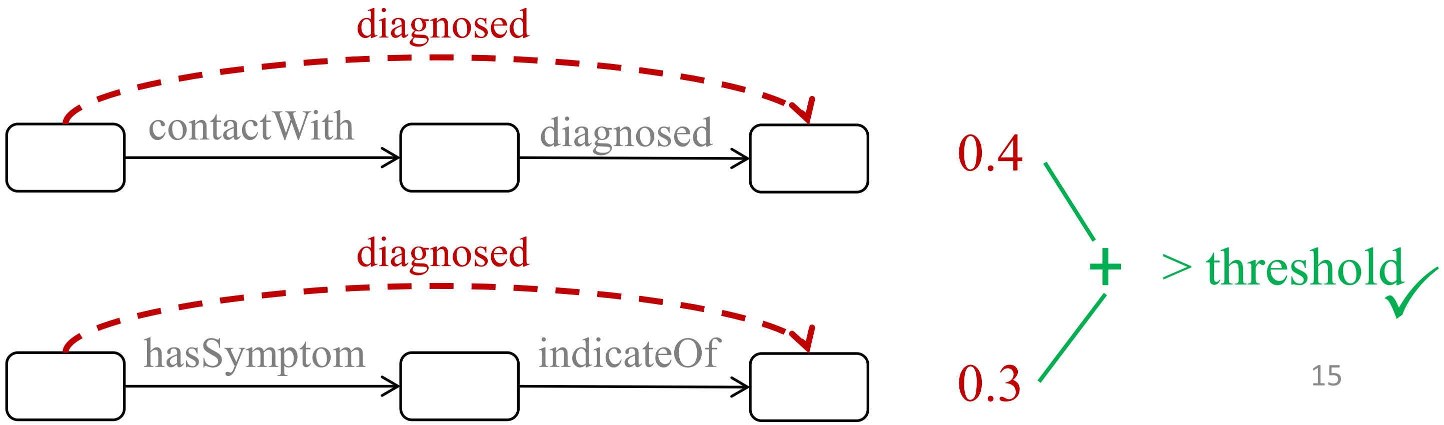


- **Question 2**: How to solve the problem and ensure faithfulness?
- We provide 2 solutions:
  1. a new rule extraction algorithm to support the "sum up" behavior
     - e.g., *diagnosed(x, y)* ← *contactWith(x, z) ∧ diagnosed(z, y)* ∧ *hasSymptom(x, v) ∧ indicateOf(v, y)*
     - by analyzing the underlying mechanism of the model, and proving the faithfulness



How does the model work?

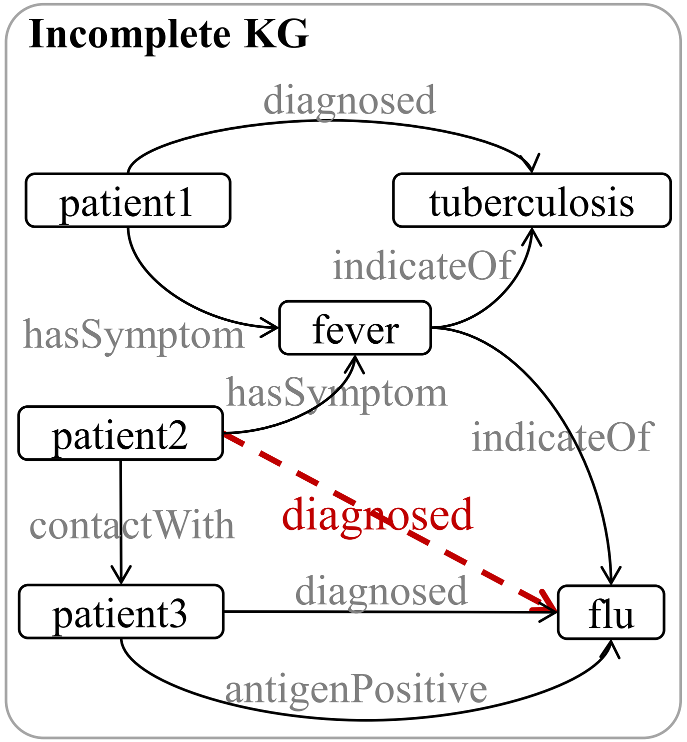
- for each relation, e.g., *diagnosed*
- it learns a weight for each "chain pattern"
- given a query *diagnosed(s, ?)*, it **sums up** all paths from *s* to the target node



Faithful Rule Learning for Knowledge Graph Completion

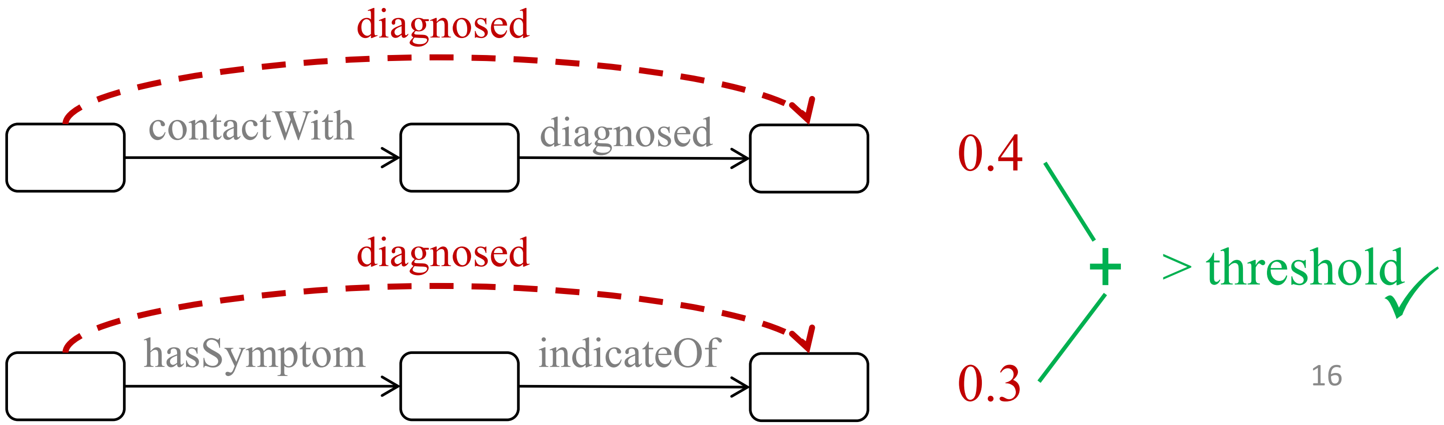
UNIVERSITY OF OXFORD

- **Question 2**: How to solve the problem and ensure faithfulness?
- We provide 2 solutions:
  1. a new rule extraction algorithm to support the "sum up" behavior
  2. a new (simplified) model with max aggregation for all paths
     - it only considers one path with max weight—can be captured by chain-like rules



How does the model work?

- for each relation, e.g., *diagnosed*
- it learns a weight for each "chain pattern"
- given a query *diagnosed(s, ?)*, it **sums up** all paths from *s* to the target node



Faithful Rule Learning for Knowledge Graph Completion

UNIVERSITY OF OXFORD

- **Question 2**: How to solve the problem and ensure faithfulness?
- We provide 2 solutions[5]:
  1. a new rule extraction algorithm to support the "sum up" behavior
  2. a new (simplified) model with max aggregation for all paths
     - it only considers one path with max weight—can be captured by chain-like rules

- **Takeaways**
  - faithful rule extraction has high computational complexity
  - the simplified model: trade-off between expressive power vs. ease of rule extraction
  - similar conclusions exist for other rule learning models, *e.g.,* NeuralILP[1,6]

[1] Yang et al. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. NeurIPS 2017

[5] Wang et al. Faithful Rule Extraction for Differentiable Rule Learning Models. ICLR 2024

[6] Tena Cucala et al. Faithful Approaches to Rule Learning. KR 2022

Faithful Rule Learning over Databases

- **Question 3**: Can we extend to **more complex** graph data?
  - from knowledge graphs to databases (hypergraphs)
  - each fact contains multiple entities

Name	City	Country
Alice	*Boston*	*US*
…	…	…

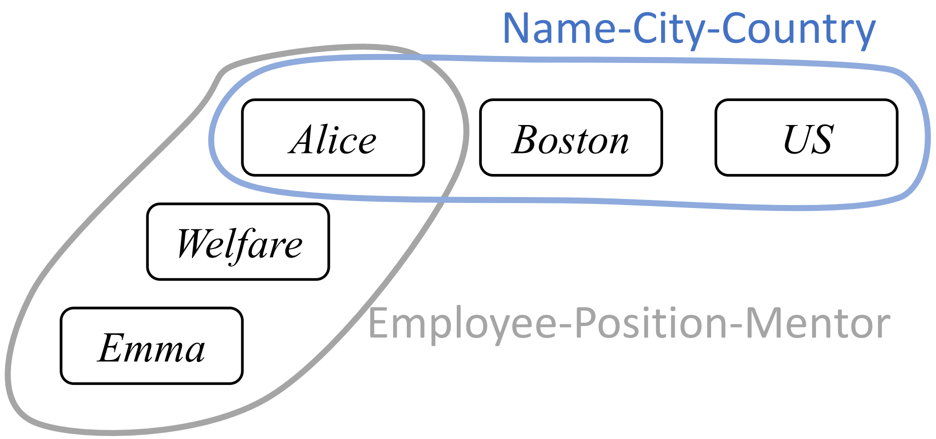
Employee	Position	Mentor
Emma	*Welfare*	*Alice*
…	…	…

Name	Org.	Country
Emma	*MIT*	*US*
…	…	…

An example database instance

Relational databases:

- each value in a row: a node (entity)
- each row (fact): a hyperedge containing multiple nodes



- **Task**: Tabular data cell completion
  - real-world tables are usually **incomplete** with missing values
  - data cell completion: to infer the missing value based on existing facts
  - given a query ***relation(a, b, ?, c, …)***, to predict the entity that matches ***?***

Name	City	Country
Alice	*Boston*	*US*
…	…	…
Emma	*Boston*	?

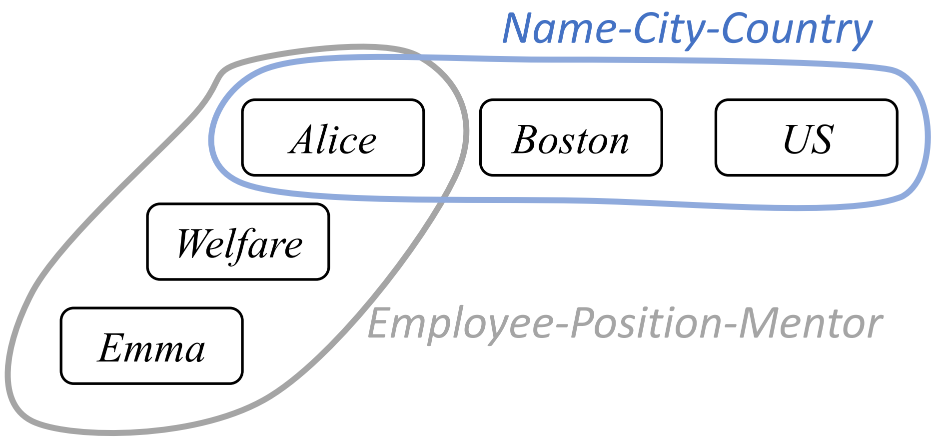
Employee	Position	Mentor
Emma	*Welfare*	*Alice*
…	…	…

Name	Org.	Country
Emma	*MIT*	*US*
…	…	…

An example database instance

Relational databases:

- each value in a row: a node (entity)
- each row (fact): a hyperedge containing multiple nodes



Faithful Rule Learning over Databases

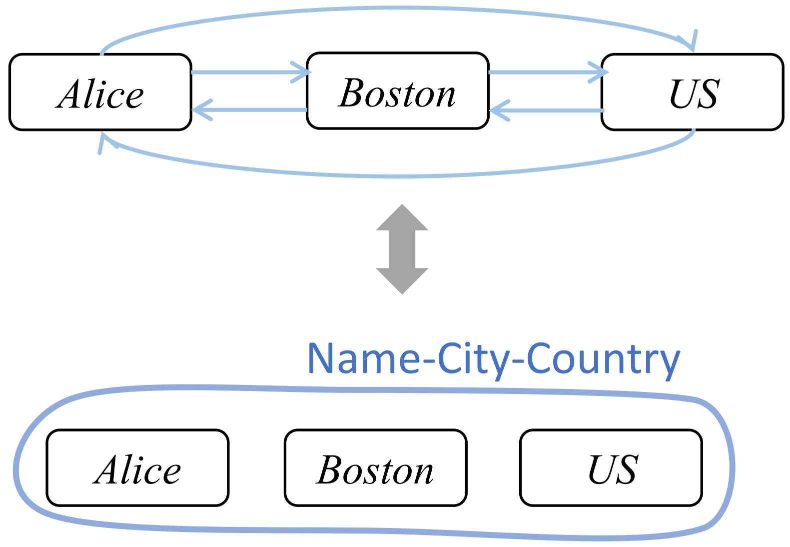
- **Task**: Tabular data cell completion
  - given a query *relation(a, b, ?, c, …)*, to predict the entity that matches *?*
- We consider binary relationships between entities

Name	City	Country
Alice	*Boston*	*US*
…	…	…
Emma	*Boston*	**?**

Employee	Position	Mentor
Emma	*Welfare*	*Alice*
…	…	…

Name	Org.	Country
Emma	*MIT*	*US*
…	…	…

An example database instance



Faithful Rule Learning over Databases

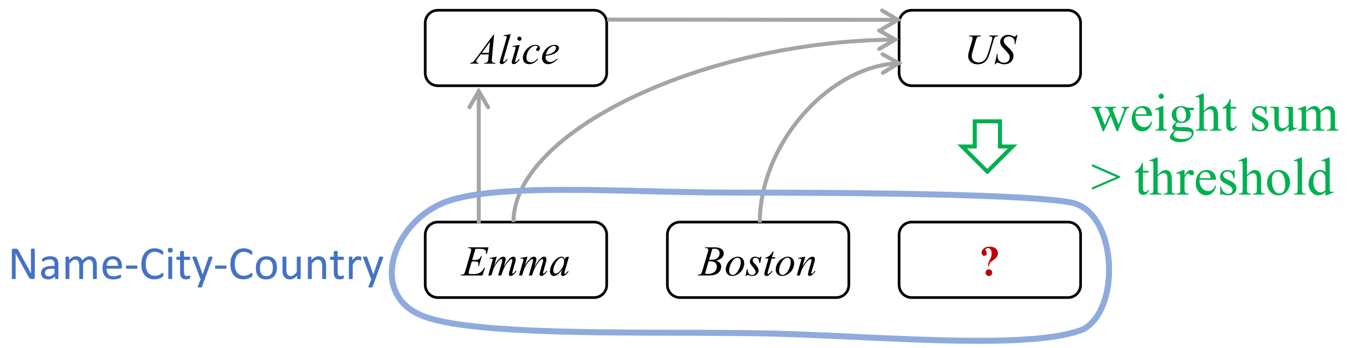
- **Task**: Tabular data cell completion
  - given a query *relation(a, b, ?, c, …)*, to predict the entity that matches *?*
- We consider binary relationships between entities
- We propose a **model**: predict missing value by aggregating paths from existing entities

Name	City	Country
Alice	*Boston*	*US*
…	…	…
Emma	*Boston*	?

Employee	Position	Mentor
Emma	*Welfare*	*Alice*
…	…	…

Name	Org.	Country
Emma	*MIT*	*US*
…	…	…

An example database instance



- aggregate paths from existing entities in the query
- weight of each path: learned by the model

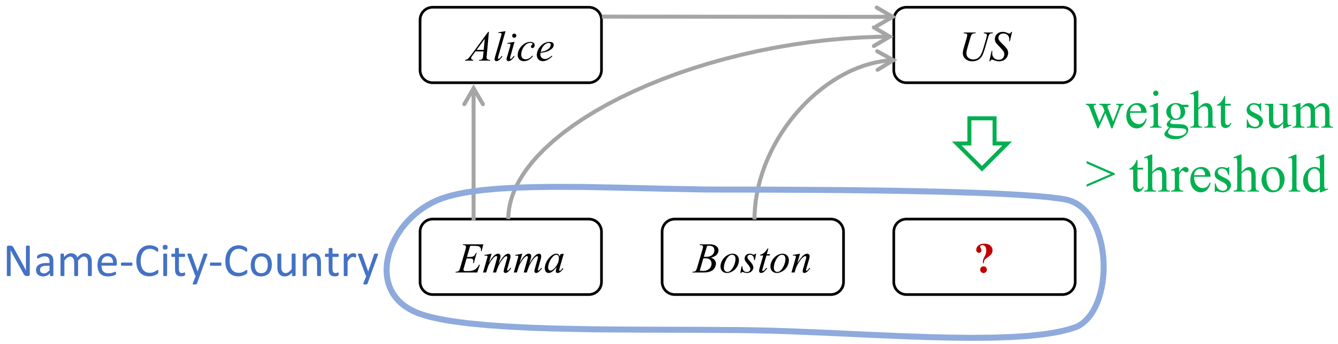
Faithful Rule Learning over Databases

UNIVERSITY OF OXFORD

- **Task**: Tabular data cell completion
  - given a query *relation(a, b, ?, c, …)*, to predict the entity that matches *?*
- We propose a **model**: predict missing value by aggregating paths from existing entities
- We propose **faithful rule extraction** algorithms for the model

- **Evaluation results**
  - SOTA performance for the task
  - better explainability

  (compared to hypergraph neural networks)



- aggregate paths from existing entities in the query
- weight of each path: learned by the model

Future Directions

- **We aim to**: enhance the explainability of ML models by **faithful rule extraction**

- **Future Directions**
  1. Diversified applications
     - *e.g.,* entity resolution by answer set programming[7], question answering
  2. Features of the extracted rule set
     - *e.g.,* rule multiplicity
  3. Practical & Reasonable benchmark generation
     - instead of using random splits for evaluation

[7] Xiang et al. ASPEN: ASP-Based System for Collective Entity Resolution. KR 2024

Thank You!

Q & A

xiaxia.wang@cs.ox.ac.uk