Link Prediction in Knowledge Graphs with Concepts of Nearest Neighbours

[presented at ESWC'19, extended in journal Data Science]

Sébastien Ferré
Team SemLIS, Univ Rennes, CNRS, IRISA

RoD (Reasoning on Data)
Symposium Madics, July 7th, 2020

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALÉATOIRES

UMR IRISA SemLIS

Introduction

- Knowledge Graphs (KG) are widely used for
  - representation (RDF)
  - reasoning (RDFS, OWL)
  - querying (SPARQL)
- KGs are often incomplete
  and completing them manually is tedious
- Inductive inference by AI means is desirable and feasible
  1. *somebody born in Milano has probably Italian nationality*
  2. *somebody speaking Spanish was probably born in Spain or Latin America*

Introduction

- Knowledge Graphs (KG) are widely used for
  - representation (RDF)
  - reasoning (RDFS, OWL)
  - querying (SPARQL)
- KGs are often incomplete
  and completing them manually is tedious
- Inductive inference by AI means is desirable and feasible
  1. *somebody born in Milano has probably Italian nationality*
  2. *somebody speaking Spanish was probably born in Spain or Latin America*

Introduction

- Knowledge Graphs (KG) are widely used for
  - representation (RDF)
  - reasoning (RDFS, OWL)
  - querying (SPARQL)
- KGs are often incomplete
  and completing them manually is tedious
- Inductive inference by AI means is desirable and feasible
  1. *somebody born in Milano has probably Italian nationality*
  2. *somebody speaking Spanish was probably born in Spain or Latin America*

Link Prediction (aka. KG Completion)

- Definition: predict the missing head or tail of a triple
  - given a KG and an incomplete triple (Pablo, bornIn, ?),
  - predict Spain, Mexico, Peru, ...
- Challenges as a learning task
  1. complex multi-relational data
  2. many relations to predict (e.g. bornIn)
  3. many possible values for each relation (e.g., all countries)
  4. multi-valued relations (e.g., actor from films to actors)
- Existing approaches
  - latent features: tensor factorization, graph embeddings, ...
    ex: *RESCAL, DistMult, TransE, HolE, ComplEx, R-GCN, ConvE*
    \+ state-of-the-art performance, – no explanations
  - observed features: paths, graph patterns, ...
    ex: *PRA (random walks), AMIE+, AnyBURL (association rules)*
    \+ (partial) explanations, – expressivity (mostly constant-free paths)
  - both: – costly learning phase

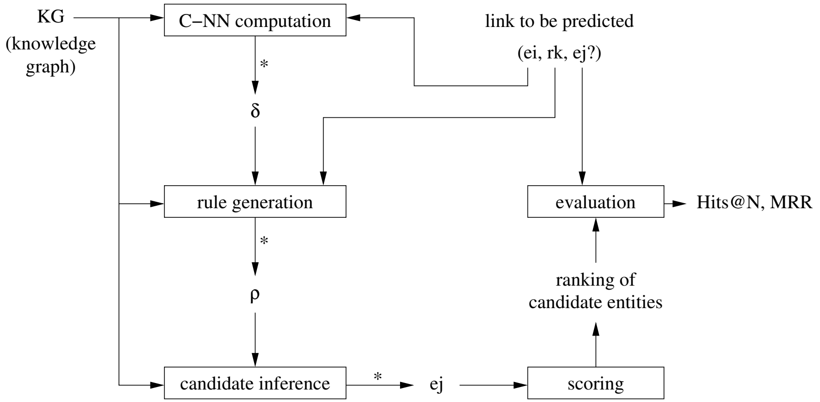
Link Prediction (aka. KG Completion)

- Definition: predict the missing head or tail of a triple
  - given a KG and an incomplete triple (Pablo, bornIn, ?),
  - predict Spain, Mexico, Peru, ...
- Challenges as a learning task
  1. complex multi-relational data
  2. many relations to predict (e.g. bornIn)
  3. many possible values for each relation (e.g., all countries)
  4. multi-valued relations (e.g., actor from films to actors)
- Existing approaches
  - latent features: tensor factorization, graph embeddings, ...
    ex: *RESCAL, DistMult, TransE, HolE, ComplEx, R-GCN, ConvE*
    + state-of-the-art performance, **–** no explanations
  - observed features: paths, graph patterns, ...
    ex: *PRA (random walks), AMIE+, AnyBURL (association rules)*
    + (partial) explanations, **–** expressivity (mostly constant-free paths)
  - both: **–** costly learning phase

We propose a new kind of approach

From k-NN to C-NN (Concepts of Nearest Neighbours)

- instance = KG entity
- distance = graph pattern shared between two entities



Overview

Knowledge Graph = Entities + Relations + Triples

female

male

Diana

Charles

Kate

William

Harry

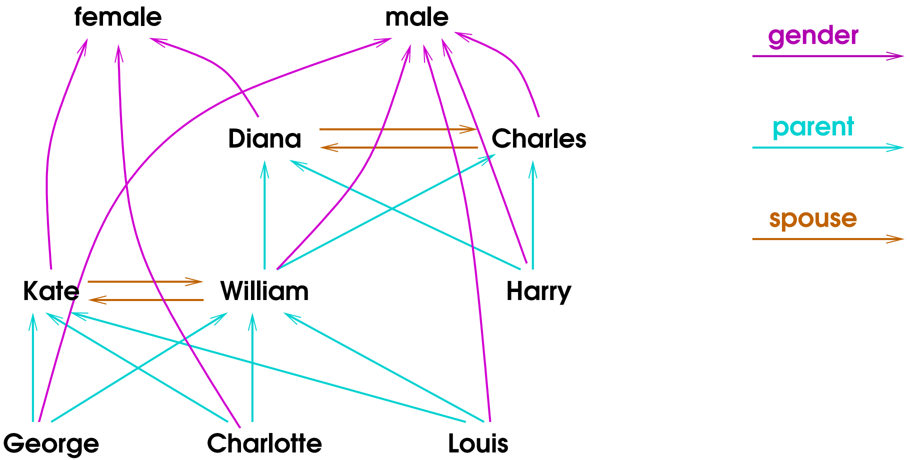
George

Charlotte

Louis

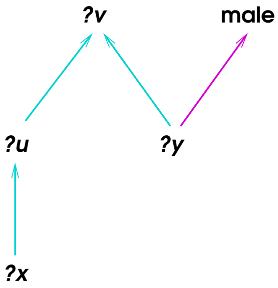
Knowledge Graph = Entities + Relations + Triples

Knowledge Graph = Entities + Relations + Triples



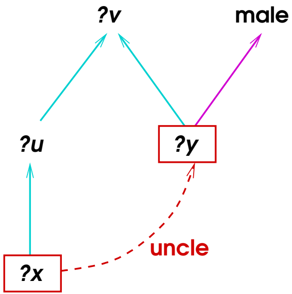
Graph Patterns, Queries, and Answers

Graph Pattern



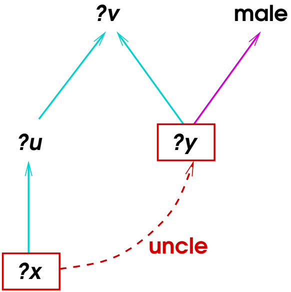
Graph Patterns, Queries, and Answers

Query



Graph Patterns, Queries, and Answers

Query



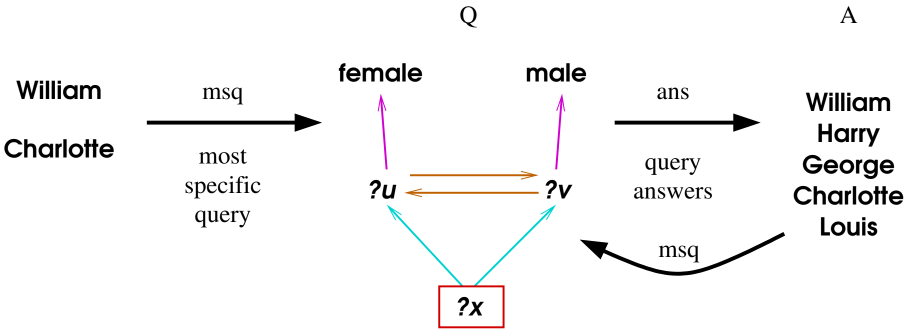
Answers (6)

x	y
George	Harry
Charlotte	Harry
Louis	Harry
George	William
Charlotte	William
Louis	William

Overview

Graph Concepts (from Graph-FCA [Ferré, 2015])

Starting from two entities:



A graph concept is a pair (A, Q), satisfying:

- $A = ans(Q)$: extension, set of concept instances
- $Q = msq(A)$: intension, concept description

Conceptual Distance

Definition

The conceptual distance between two entities e_i, e_j is defined as the most specific graph concept that contains them:
$\delta(e_i, e_j) = (A_{ij}, Q_{ij})$ where

- $Q_{ij} = msq(\{e_i, e_j\})$: what they have in common
- $A_{ij} = ans(Q_{ij})$: which entities range between them

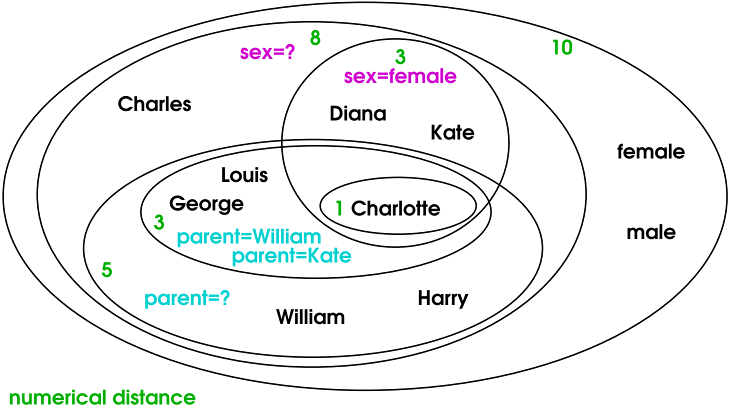
- δ is a symbolic distance
  - distances are partially ordered (concept inclusion)
- δ verifies distance axioms (positivity, symmetry, triangular ineq.)
  - with bottom concept as zero
  - with concept union as addition
- numerical measures can be derived
  - $dist(e_i, e_j) = |ext(\delta(e_i, e_j))|$: distance as number of answers
  - $sim(e_i, e_j) = |int(\delta(e_i, e_j))|$: similarity as size of the query

Concepts of Nearest Neighbours (C-NN)

Given a knowledge graph $K = (E, R, T)$, and an entity $e \in E$:

$$CNN(e, K) = \{\delta(e, e') \mid e' \in E\}$$

Example for $e = Charlotte$ (6 C-NNs)



Overview

Generation of Inference Rules from C-NNs

- Compared to k-NN classification
  - concepts of neighbours instead of neighbour instances
  - inference rules instead of labelling class
- For each $\delta = (A, Q) \in CNN(e_i)$, where $Q = x \leftarrow P$
- Two kinds of rules are generated for the target relation r_k
  1. **by-copy rules**: $P \rightarrow (x, r_k, e_j)$ for $e_j \in range(r_k)$
     - if somebody was born in Spain, then she probably speaks Spanish
     - inferred entities: $\{e_j\}$

$$conf := \frac{|ans(x \leftarrow P, (x, r_k, e_j))|}{|ans(x \leftarrow P)|}$$

  2. **by-analogy rules**: $P \rightarrow (x, r_k, y)$ for $y \in Vars(P), y \neq x$
     - e_i is to e_j as x is to y in P (analogical proportion)
     - if somebody has a father whose wife is Y, then she probably has Y as a mother
     - inferred entities: $ans(y \leftarrow P, (x = e_i))$

$$conf = \frac{|ans((x, y) \leftarrow P, (x, r_k, y))|}{|ans((x, y) \leftarrow P)|}$$

Generation of Inference Rules from C-NNs

- Compared to k-NN classification
  - concepts of neighbours instead of neighbour instances
  - inference rules instead of labelling class
- For each $\delta = (A, Q) \in CNN(e_i)$, where $Q = x \leftarrow P$
- Two kinds of rules are generated for the target relation r_k
  1. **by-copy rules**: $P \rightarrow (x, r_k, e_j)$ for $e_j \in range(r_k)$
     - if somebody was born in Spain, then she probably speaks Spanish
     - inferred entities: $\{e_j\}$

$$conf := \frac{|ans(x \leftarrow P, (x, r_k, e_j))|}{|ans(x \leftarrow P)|}$$

  2. **by-analogy rules**: $P \rightarrow (x, r_k, y)$ for $y \in Vars(P), y \neq x$
     - e_i is to e_j as x is to y in P (analogical proportion)
     - if somebody has a father whose wife is Y, then she probably has Y as a mother
     - inferred entities: $ans(y \leftarrow P, (x = e_i))$

$$conf = \frac{|ans((x, y) \leftarrow P, (x, r_k, y))|}{|ans((x, y) \leftarrow P)|}$$

Scoring and Ranking Inferred Entities

Maximum Confidence (introduced for AnyBURL [Meilicke, 2019])

- The score of each inferred entity e_j is
  - the list of rule confidence measures (above 0.01)
  - in decreasing order
  - from all rules inferring e_j
  - ex: 0.94 0.86 0.33 ...
- Ranking of all inferred entities
  - in decreasing lexicographic order

e1	0.94	0.86	0.33 ...
e2	0.94	0.86	
e3	0.94	0.67	0.43 ...
e4	0.55	0.43	0.33 ...
...	...		

Overview

Example of correct inferences

On the Mondial dataset, with timeout = 0.1+0.1s

1. mountain "Reuss" $\xrightarrow{\text{is in mountain range}}$ Alps (0.50 0.38 0.27)
   28 C-NNs, best explanation: *located in a place speaking Italian and German* (by-copy rule)
2. mountain "Matterhorn" $\xrightarrow{\text{is located in}}$ Switzerland (0.42 0.36 0.29)
   30 C-NNs, best explanation: *two mountains in the same range tend to have the same location* (by-analogy rule)
3. river "Araguaia" $\xrightarrow{\text{is located in}}$ Brazil (0.93 0.92 0.92)
   47 C-NNs, best explanation: *a water shares location with an island it contains* (by-analogy rule)

Experimental Results: WordNet Benchmarks

Approach	WN18 H@1	WN18 H@10	WN18 MRR	WN18RR H@1	WN18RR H@10	WN18RR MRR
Freq	1.8	5.0	2.9	1.5	4.4	2.6
Latent-based						
DISTMULT	70.1	94.3	81.3	-	-	-
ANALOGY	93.9	-	94.2	-	-	-
KB_LR	-	95.1	93.6	-	-	-
R-GCN+	69.7	96.4	81.9	-	-	-
ConvE	93.5	95.5	94.2	39.0	48.0	46.0
ComplEx-N3	-	96.0	95.0	-	**57.0**	**48.0**
CrossE	74.1	95.0	83.0	-	-	-
Rule-based						
AMIE+	87.2	94.8	-	35.8	38.8	-
RuleN	94.5	95.8	-	42.7	53.6	-
AnyBURL	93.9	95.6	95.0	**44.6**	55.5	**48.0**
C-NN (ours)	**96.7**	**97.2**	**96.9**	44.4	51.9	46.9
C-NN – best other	+2.2	+0.8	+1.9	-0.2	-5.1	-1.1
C-NN – best rule-based	+2.2	+1.4	+1.9	-0.2	-3.6	-1.1

Experimental Results: Freebase Benchmarks

Approach	FB15k H@1	FB15k H@10	FB15k MRR	FB15k-237 H@1	FB15k-237 H@10	FB15k-237 MRR
Freq	14.3	28.5	19.2	17.5	35.6	23.6
Latent-based						
DISTMULT	52.2	81.4	63.4	10.6	37.6	19.1
ANALOGY	64.6	-	72.5	-	-	-
KB_LR	74.2	87.3	79.0	22.0	48.2	30.6
R-GCN+	60.1	84.2	69.6	15.1	41.7	24.9
ConvE	67.0	87.3	74.5	**23.9**	49.1	31.6
ComplEx-N3	-	**91.0**	**86.0**	-	**56.0**	**37.0**
CrossE	63.4	87.5	72.8	21.1	47.4	29.9
Rule-based						
AMIE+	64.7	85.8	-	17.4	40.9	-
RuleN	77.2	87.0	-	18.2	42.0	-
AnyBURL	80.4	89.0	83.0	23.0	47.9	30.0
C-NN (ours)	**82.7**	89.0	84.9	22.2	44.6	29.6
C-NN – best other	+2.3	-2.0	-1.1	-1.7	-11.4	-7.4
C-NN – best rule-based	+2.3	0.0	+1.9	-0.8	-3.3	-0.4

Overview

Conclusion

We have proposed a symbolic approach to link prediction that

- is competitive with both rule-based and latent-based approaches
- provides explanations for each inference
  - local explainability (not global explainability)
- avoids the training phase (instance-based learning)
  - which enables application to dynamic KG
- has a controllable runtime (anytime algorithm)
  - timeout is the only significant hyperparameter

In short

An adaptation of k-NN classification to knowledge graphs with conceptual distances.

Perspectives

- To extend graph patterns
  ex: inequalities; richer filters on numbers, strings, dates
- To optimize the computation of C-NNs
  ex: partitioning strategies, parallelization
- To explore other kinds of inference
  ex: structured prediction, ...
- To evaluate on other tasks
  ex. relation extraction

The End

Thanks for listening !

Algorithmic and Practical Aspects

[see ESWC'18 paper on approximate query answering]

- $CNN(e, K)$ are computed by incrementally partitioning E
  - triples describing e are used as discriminating features
  - PRO: the number of clusters is bounded by $|E|$
- the partitioning algorithm is anytime
  - only coarser partition if stopped before completion
- previous experiments have shown greater efficiency compared to
  - computing conceptual distances with each entity
  - applying query relaxation to the description of e