



Completeness, Recall, and Negation in Open-World Knowledge Bases

Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian Suchanek

- 1. Introduction & Foundations (Simon)
- 2. Predictive recall assessment (Fabian)
- 3. Counts from text and KB (Shrestha)
- 4. Negation (Hiba)
- 5. Wrap-up (Simon)

Open-world Assumption

Wikidata - Awards won by Stephen Hawking

42 awards (in KB)



Problem

Wikidata - Awards won by Stephen Hawking

42 awards (in KB)



Existing positive-only KBs are <u>unaware</u> of negation.

Solution: materialize ALL negations?

Wikidata - Awards won by Stephen Hawking

42 won awards, 30000 possible awards



Challenge1: Set of negative statements is quasi-infinite! Challenge2: Not every <u>missing</u> link is a negation

Do knowledge bases allow negatives?

- ERDF knowledge bases
 - Extended RDF to include \sim and \neg
 - In OWA ~awardRecevied(Hawking, Nobel in Physics)
- Count properties
- Object = No-value
- Negated properties
 - DBpedia, *never exceed alt*
 - Knowlife, not caused by
- Deprecated rank
 - Flagging triples known to include errors

Good news: formally defined negative information **Bad news:** KB Completion challenges, limited domains, no active collection

Analyti et al., "Stable Model Theory for Extended RDF Ontologies", ISWC'05



Research Problem



¬ (award, Nobel Prize in Physics)





(award, Oscar)

¬ (headquarters, Silicon Valley)



Methodologies

★ Statistical Inferences

Textual Extractions

Language Models

Method: Peer-based Negation Inference

Given entity e in KB:

■Gist:

- select highly related entities (peers)
- make local CWA (within peer group)
- infer <u>expectations</u> about e (candidate negations)
- validate and score expectations
- Output: top correct and salient negations about e.

8

Collect Peers

What is a similar entity?

Class-based

Stephen Hawking: Physicist Albert Einstein, Marie Curie, Max Planck

Jaccard-similarity

predicate-object pairs shared by entities: Hawking AND Einstein = 423/750 same profession, education, memberships, awards,...

Embedding-based similarity

low-dimensional latent representations **cosine_sim(***embeddings*(*Hawking*), *embeddings*(*Jane Wilde*))



Accurate

9

Retrieve candidate negations



Validate candidates

Challenge: correctness of inferred negations. Are they <u>negatives</u> or <u>missing facts</u>?

(award, Nobel Prize in Physics) (sibling, Edward Hawking) (award, Copley Medal) (hobby, sailing) (sibling, Maja Einstein)

Retain candidate only in presence of other values (i.e., PCA) (Hawking, award, {Copley Medal, …}) ⊨ ¬ (award, Nobel Prize in Physics) (Hawking, hobby, Ø) ⊭ ¬ (sailing, reading)

Significantly boosts correctness of deductions: by 30%.

Validate candidates

Challenge: correctness of inferred negations. Are they <u>negatives</u> or <u>missing facts</u>?

entity (award, Nobel Prize in Physics) (sibling, Edward Hawking) (award, Copley Medal) (hobby, sailing) (sibling, Maja Einstein)

Retain candidate only in presence of other values (i.e., PCA) (Hawking, award, {Copley Medal, …}) ⊨ ¬ (award, Nobel Prize in Physics) (Hawking, hobby, Ø) ⊭ ¬ (sailing, reading)

Significantly boosts correctness of deductions: by 30%. % truly negative

Rank candidates

entity



Learn2Rank:

- A. Obtain annotator judgments for statement interestingness [0..1] Is it interesting that Hawking <u>never</u> received a Nobel in Physics? .. is <u>not</u> the sibling of Maja Einstein?
- B. Train supervised model: Linear Regression feature: peer frequency, entity and predicate authority, word embeddings
- C. Rank assertions by descending scores

Methodologies

Statistical Inferences

Textual Extractions

Language Models

Method: Mine Negations from User Query Logs

• Wisdom of the crowd: Search-engine autocompletion provides access to salient user assertions

- Probing with negated prefixes
 - Why didn't <e>
 - Why hasn't <e>
 - Why wasn't <e>
 - ...



- why didn't stephen hawking
- Q why didn't stephen hawking **get a nobel prize**
- Q why didn't stephen hawking die
- Q why didn't stephen hawking get a knighthood
- Q why isn't Switzerland
- Q why isn't switzerland in the eu
- Q why isn't switzerland part of germany
- Q why isn't switzerland in nato

Method: Mine Text Revisions

- Anti-knowledge base (AKB) KB of common factual mistakes Complement the positive-only KB
- Gist:

Exploit entity/number swaps in Wikipedia update logs Score using web hits



Revision 505 Einstein was born in Vienna.

Revision 506 Einstein was born in Ulm.

Karagiannis et al., "<u>Mining an "Anti-Knowledge Base</u>" from Wikipedia Updates with Applications to Fact <u>Checking and Beyond</u>", PVLDB'19 **Methodologies**

Statistical Inferences

Textual Extractions

★ Language Models

Method: Unsupervised Discovery of Negatives in Commonsense KBs

- NegatER, steps:
- 1) Fine-tuning BERT using positives commonsense knowledge.
- 2) Generate corruptions of KB statements plausible candidate negatives by corrupting positives
- 3) Estimate contradiction score over fine-tuned BERT

(horse, IsA, expensive pet) (cat, IsA, expensive pet) (goldfish, IsA, expensive pet) (horse, IsA, expensive car)

ConceptNet

18

Overview of methods

Research problem:

Given an entity, compile a list of <u>correct</u> & <u>salient</u> negations.

Paper	Method	Advantages	Limitations
AKBC'20 Arnaout et al.	Peer-based	Salience Coverage Canonicalization	Correctness
EMNLP'21 Safavi et al.	Language Models	Salience Coverage	Correctness Canonicalization
PVLDB'19 Karagiannis et al.	Anti-KB (mining revisions)	Correctness	Salience Coverage Canonicalization
AKBC'20 Arnaout et al.	Query-logs (pattern-based)	Salience Correctness	Coverage Canonicalization

Next: existing tools and datasets about useful negations.

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

Anti-KB (dataset)

Ranked common factual mistakes from Wikipedia

Explainable salient negations (dataset) "" did not know that!"

Order oriented peer-based negation inferences.



Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

Anti-KB (dataset)

Ranked common factual mistakes from Wikipedia

Explainable salient negations (dataset) "I did not know that!"

Order oriented peer-based negation inferences.



Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted



- Peer-based method.
- Interesting negations about 500K entities.

ome Documentation	Search by statement Conta	act		
	Negative statements.		Albert Einstein	Compared with
	doctoral student:	none.	13 🕢 0 🗶 6 🔘 0 🔘	
Albert Einste Go!	Click here for a possible answer. Positive for: Max Planck, Wolfgang P	auli, (6) more		Max Planck
 Live SPARQL validation Pre-computed validation 	member of:	-Russian Academy of Sciences.		Erwin Schrödinger
isplay: All statements ✓	True Values: Royal Society; French A Positive for: Max Planck, Erwin Schrö	cademy of Sciences; (8) more bdinger, (2) more		?
imilarity function:				Wolfgang Dauli
Wikipedia embeddin 🗸 legation type:	award received:	-Fellow of the American Physical Society.		wongang Pauli
Regular (no lifting) 🗸 lumber of statements:	True Values: Matteucci Medal; New . Positive for: Erwin Schrödinger. Rich	lersey Hall of Fame; (8) more ard Fevnman. (1) more		Niels Bohr
k=3 ∽				
WIKINGATA		Read more		

Arnaout et al., "Wikinegata: A Knowledge Base with Interesting Negative Statements", VLDB'21

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

★ Anti-KB (dataset) 🕻

Ranked common factual mistakes from Wikipedia

Explainable salient negations (dataset) ""I did not know that!"

Order oriented peer-based negation inferences.



Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted

Anti-knowledge base

Data available upon request

• Wikipedia update log.



• <u>116k</u> likely mistakes where entities or numbers were corrected.

Penicillin was discovered in 1928 by Scottish scientist Alexander Baldwin.

Alexander Flemming.



Confidence score = 0.619

Karagiannis et al., "<u>Mining an "Anti-Knowledge Base" from Wikipedia Updates with Applications to Fact</u>₂₄ <u>Checking and Beyond</u>", PVLDB'19

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

Anti-KB (dataset)

Ranked common factual mistakes from Wikipedia

★ Explainable salient negations (dataset) "I did not know that!"

Order oriented peer-based negation inferences.



Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted

"I did not know that!"

- Dataset with <u>12.5M</u> salient negations
- Inferred using the peer-based method (with ordered peers).

Entity	Negation	Contextualized Verbalization
Abraham Lincoln	¬(death cause, natural)	Unlike the previous 17 U.S. presidents.
Jeff Bezos	¬(occupation, politician)	Unlike the previous 17 of 21 Time Person of the Year winners.
Angela Merkel	¬(gender, male)	Unlike the previous 6 <u>Chairmen of the CDU</u> .

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

Anti-KB (dataset)

Ranked common factual mistakes from Wikipedia

Explainable salient negations (dataset) "I did not know that!"

Order oriented peer-based negation inferences.

★ ANION (dataset) 🚺

Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted

ANION

 Commonsense knowledge graph 624K if-then rules

https://github.com/liweijiang/anion



28

Resources

Wikinegata (online platform)

Browse interesting negations about Wikidata entities

Anti-KB (dataset)



Ranked common factual mistakes from Wikipedia

Explainable salient negations (dataset) ""I did not know that!"

Order oriented peer-based negation inferences.



Commonsense KB focusing on negated events

Google Hotel Search (online platform)



Hotel booking with negative features asserted



Data crawled from:

- Hotel websites
- Third-party services
- User reviews



https://www.google.com/travel/hotels/



Takeaway: negation

- KBs lack meaningful negative knowledge
- Interest in the *explicit addition* of negation to OWKBs
- Applications:
 - Commercial decision making(e.g., hotel booking)
 - General-domain QA systems (e.g., is Switzerland a member of the EU?)
- Methodologies: statistical inferences, text extractions, language models
- KB Challenges:
 - <u>Class hierarchies</u> profession not *producer* but *film-producer*
 - <u>Modelling issues</u> field not *Information Technology* but *Informatics*
 - o <u>Maintenance</u>

¬(Elon Musk, award, Time's Person of the Year) valid negation up until 2021