Knowledge Graph Simple Question Answering for Unseen Domains

Georgios Sidiropoulos, Nikos Voskarides, Evangelos Kanoulas

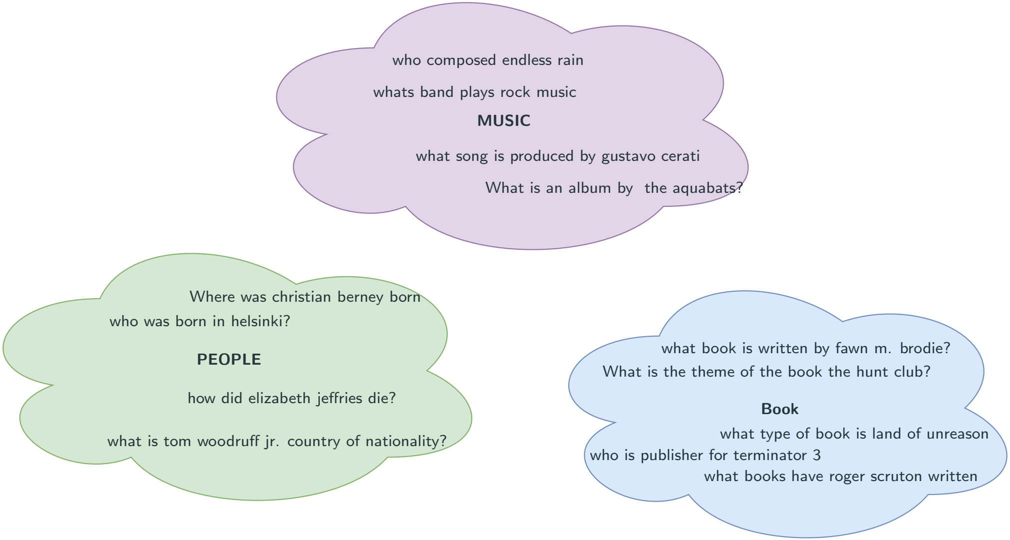
University of Amsterdam

Introduction

Knowledge Graph Simple Question Answering

Knowledge Graph Simple Question Answering (KGSQA):
Automatically answer natural language questions using a single fact contained in a KG.

Knowledge Graph Simple Question Answering



Knowledge Graph Simple Question Answering for Unseen Domains

Why is it essential to study KGSQA for unseen domains?

Knowledge Graph Simple Question Answering for Unseen Domains

Why is it essential to study KGSQA for unseen domains?

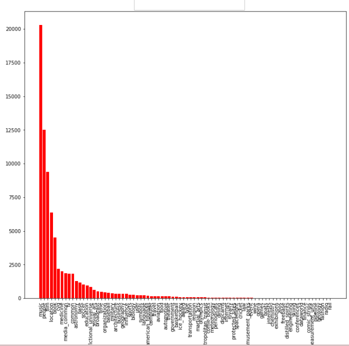
- KGs are dynamic.
  - New domains can be added introducing new entities, relations and facts.
  - No available questions for the new facts.

Knowledge Graph Simple Question Answering for Unseen Domains

Why is it essential to study KGSQA for unseen domains?

- KGs are dynamic.
  - New domains can be added introducing new entities, relations and facts.
  - No available questions for the new facts.

- The majority of data are distributed among a small number of domains.



Objective

Train a model in source domain (training set) that will perform well on target domain (test set), given the fact that source and target have been generated from different distributions.

Methodology

Pipeline

KGSQA task can be modularized into three sub-tasks (Ture and Jojic (2016)):

- Mention Detection (MD)
  - Detect the single entity mention in question.
  - Sequence tagging problem.
- Candidate Generation (CG)
  - Link the mention to an actual entity node in the KG.
  - Simple heuristics.
- Relation Prediction (RP)
  - Classify question into one of the relation types.
  - Ranking approach.

Applying model to unseen Domains

- MD, CG can scale well to unseen domains.
- RP has limited generalization ability:
    - Same relation can occur with various lexicalizations ($place\_of\_birth$ "took his first breath" or "was born").

Solution: Generate Synthetic Questions for the Target Domain

- Create synthetic questions originating from the unseen target domain.
- Use synthetic questions to assists the training of RP.
- Modify the proposed approach by Elsahar et al. (2018), for zero-shot question generation from KG.

Knowledge Graph Question Generation (KGQG)

Input: A fact from the KG; A textual context for every part of the fact.

- e.g. ("Johnny Cash", $people.cause\_of\_death.people$, "Diabetes") & ("musical artist", "death by", "disease")

Output: A synthetic question.

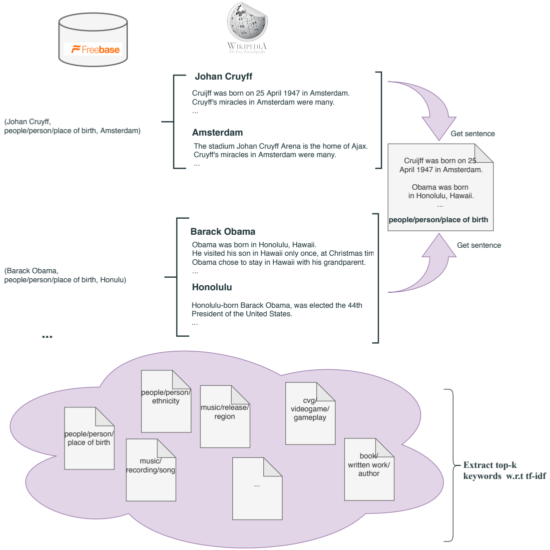
- e.g. Which disease the death of Johnny Cash?

KGQG: Relation's Textual Context

Relation's textual context:

- Noisy in many cases.
- Does not provide enough lexicalizations (limiting the variety of the generated questions).

KGQG: Relation's Textual Context



KGQG: Relation's Textual Evidence

Table 1: Examples of relation textual contexts generated using our keyword extraction approach.

Relation	Textual Context
$music.artist.label$	records, artists, album, released, label, signed, band, groups, albums, musical
$people.deceased\_person.place\_of\_death$	died, death, deaths, born, age, people, male, actors, buried, cemetery
$film.film.directed\_by$	film, director, directed, films, short, directing, producer, feature, best, award

Results

Results

Table 2: Relation Prediction accuracy w.r.t. different ways of generating synthetic training data for the unseen domain.

Synthetic training data	Macro-avg. Accuracy (%)	Micro-avg. Accuracy (%)
-	30.21	29.06
Wiki-raw-sentences	37.89	36.51
KGQG (Elsahar et al.)	67.52	69.78
KGQG (Ours)	**69.86**	**70.95**

Results

Table 3: End-to-end accuracy on the KGSQA task.

Synthetic Training Data	Relation Prediction	Macro-avg. Accuracy (%)	Micro-avg. Accuracy (%)
KGQG (Elsahar et al.)	BiLSTM (Petrochuk and Zettlemoyer)	55.49	55.11
	HR-BiLSTM (Yu et al.)	60.20	62.77
	Ours	63.90	65.18
KGQG (Ours)	Ours	**66.49**	**66.64**
Gold Questions	Ours	84.56	82.87

Conclusions

Conclusions

- Propose a framework for KGSQA that is applicable to unseen domains.
- Improve KGQG by effectively using textual information to model relations of the unseen domains.

Thanks!

References i

H. Elsahar, C. Gravier, and F. Laforest. Zero-shot question generation from knowledge graphs for unseen predicates and entity types. *arXiv preprint arXiv:1802.06842*, 2018.

M. Petrochuk and L. S. Zettlemoyer. Simplequestions nearly solved: A new upperbound and baseline approach. In *EMNLP*, 2018.

F. Ture and O. Jojic. No need to pay attention: Simple recurrent neural networks work!(for answering" simple" questions). *arXiv preprint arXiv:1606.05029*, 2016.

M. Yu, W. Yin, K. S. Hasan, C. N. dos Santos, B. Xiang, and B. Zhou. Improved neural relation detection for knowledge base question answering. In *ACL*, 2017.