Data Sharing: Querying and Linking Distributed and Autonomous Data

WP6: Intelligent Query Answering

Kwabena Nuamah, Alan Bundy

k.nuamah@ed.ac.uk, a.bundy@ed.ac.uk

Deep Algorithmic Question Answering

Motivation

Ability to reason in a step-by-step "algorithmic" manner that can be inspected and verified for its correctness in the domain of question answering (QA).

We propose ***Deep Algorithmic Question Answering** [1]*, an approach to algorithm reasoning for QA based on three desirable properties: interpretability, generalizability, and robustness. We conclude that they are best achieved with a combination of hybrid and compositional AI.

Problem

- Tasks such as the automatic selection of KBs and relevant knowledge, choice of inference algorithms, and how to combine them, are all important to fully automate the QA process.
- We argue that these tasks should be part of the AI models which are built for QA tasks, as they are key ingredients in the full automation of the QA process.

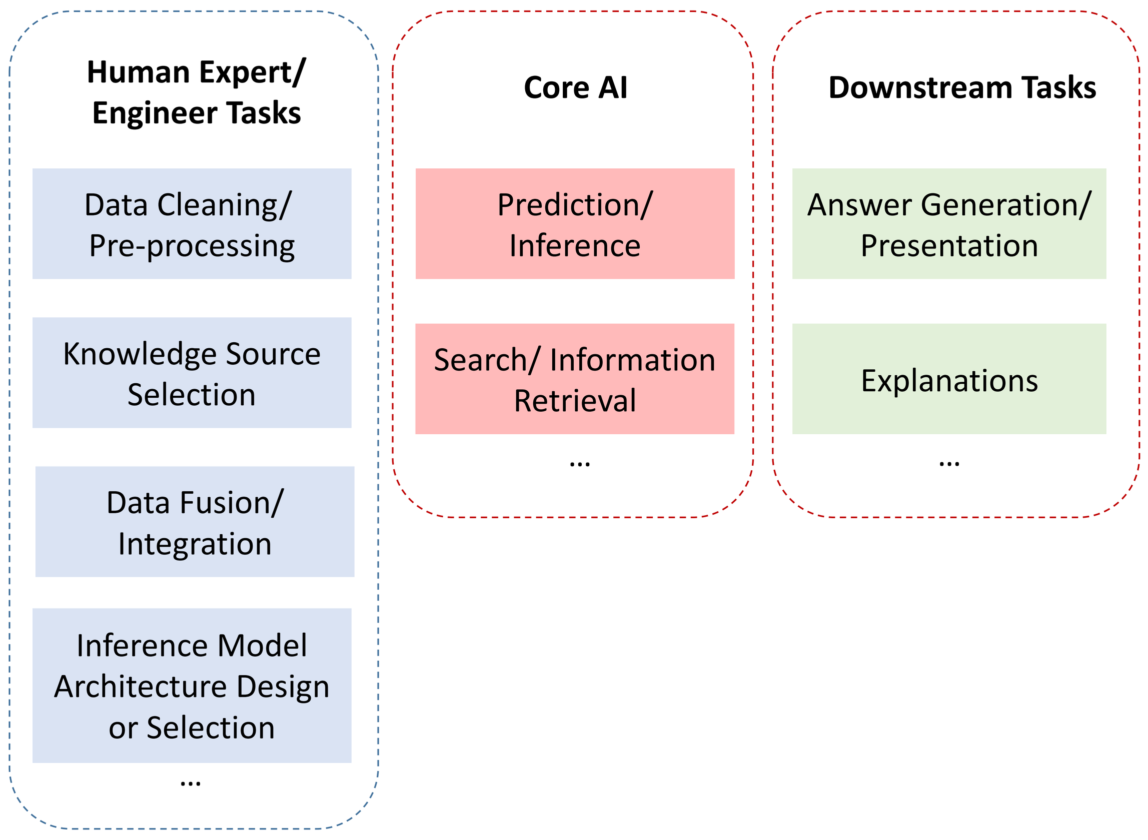
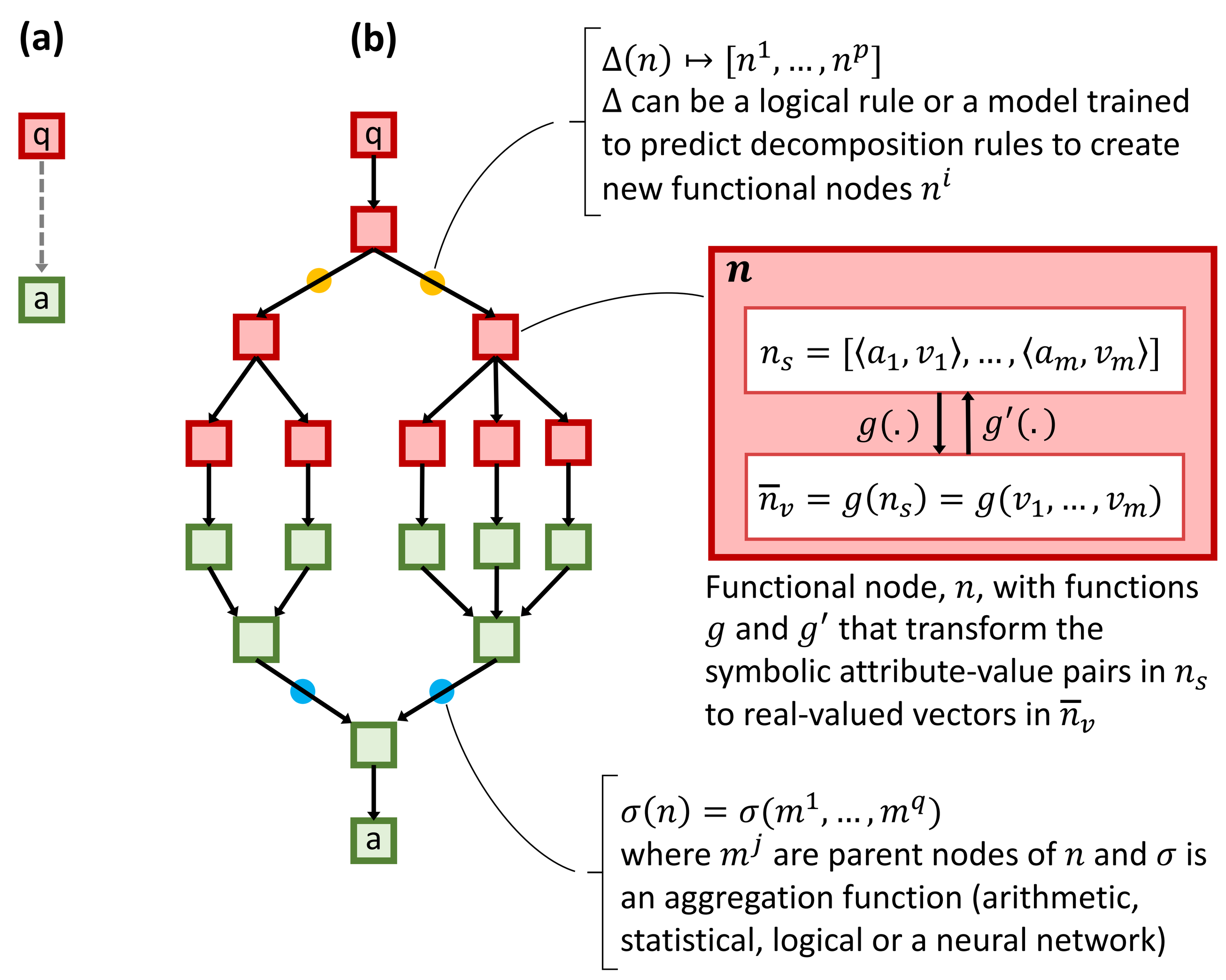


Fig. 1. *Diverse tasks that are part of the open-domain question answering process. However, most of the attention in work related to QA focus on the core AI tasks related to information retrieval, inference or prediction*



Proposed Model

Hybrid inference graphs with functional nodes.

Inference Graph

Constructed and expanded dynamically through the decompositions of its functional nodes using rules that are learned (see Fig. 2).

Functional nodes

- Represent data
- Specify operations to be applied
- Encode a model to convert between the symbolic and vectorized representations of the node.

Fig. 2.
(a) *Shows the base inference graph with a question node and an answer node that is to be inferred. They are linked by an edge that can be split by applying decomposition operations on the question node.*
(b) *An inference graph made up of functional nodes and edges labelled by operations for predicting decomposition and aggregation functions. Decomposition sub-graph (in red) is guided by a function Δ that decomposes a functional node to create new continuations of the inference graph, and aggregation sub-graph (in green) which uses a model σ to select appropriate functions to combine nodes. Functional nodes provide both a symbolic and vector representation of the node's attribute-value internal representation, as well as function g and g' for converting between the two representations.*

A Systems Approach

Improving the inference capabilities and explainability of QA systems via "whole system reasoning" [1].

Unified Decomposition & Aggregation (UDA) Rules [2]

A family of axiom schemata that are instantiated at run-time to add new axioms to a logical theory. These new axioms are implications, whose preconditions will be constructed from an analysis of the goal to be proved and the theory in which it is to be proved.

- An implication of the form:

$$Decompose(\mathcal{A}(\vec{x}), \tau) = [\mathcal{A}_j(?\vec{x}_j) | 1 \le j \le m]$$
$$\implies \mathcal{A}(\vec{h_\tau}(\epsilon ?\vec{x}_1.\ \mathcal{A}_1(?\vec{x}_1), \dots, \epsilon ?\vec{x}_m.\ \mathcal{A}_m(?\vec{x}_m)))$$

- $\mathcal{A}$ are alists, a set of attribute value pairs.
- τ is the type of decomposition and the Hilbert Epsilon operation , ϵ, converts alists from relations to functions.

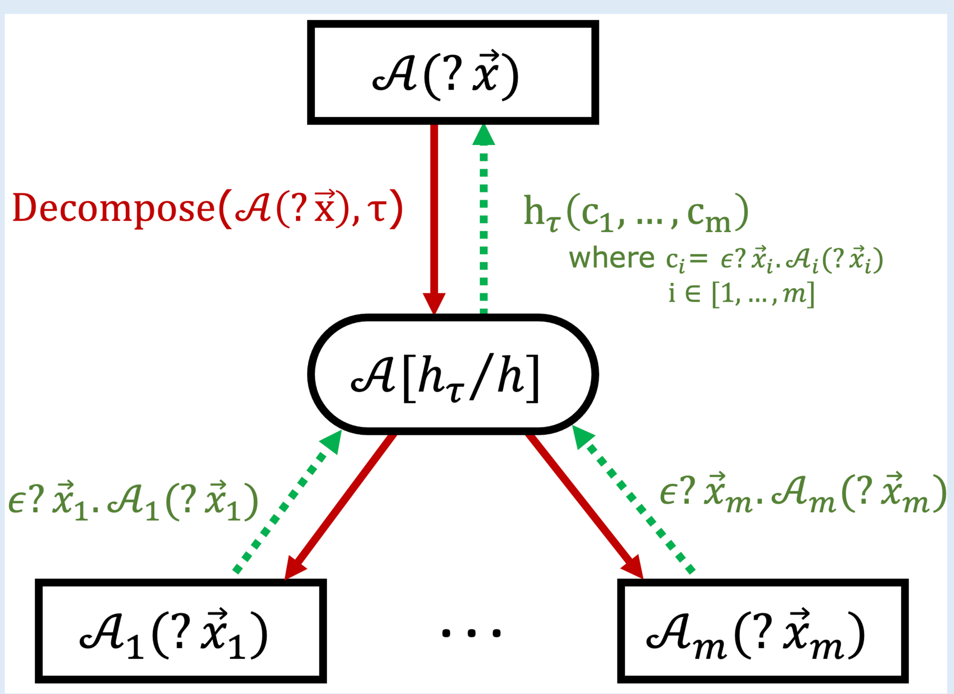


Fig. 3. *A graphical representation of a UDA rule. It shows a single step of inference in FRANK and is applied recursively to construct inference graphs. Solid red arrows show the direction of decomposition while the dashed green ones show the direction of aggregation. Instantiations (c_i) of projection variables ($?x_i$) are propagated from child alists $\mathcal{A}_i$ to their parent $\mathcal{A}$ using aggregation function h_τ of $\mathcal{A}$.*

Automatic Knowledge Source Selection [3]

Aim:
- Discover new knowledge sources.
- Identify and align equivalent entities and relationships (properties) across different knowledge graphs (KGs)

Process:
- Discovery: Crawl websites following Linked Data URIs (in Schema.org or JSON-LD formats)
- Introspection: "Upper Ontology" to capture metadata about KGs.
- Alignment: Update existing upper ontology

Usage:
- LOOKUP operation uses upper ontology to find KGs that have relevant data.

Automatic Statistical Model Selection

GPy-ABCD [4]**:**
- An iterative modular Gaussian Process regression framework
- A flexible class of nonparametric models to fit data
- Produces short text descriptions of fit models

SMART: Statistical Methodology Advisor at Reasoning Time
- Selects and performs statistical methods given a query and data features;
- Uses an ontology of various query tags, statistical methods and output types.

Acknowledgement: This work was funded by Huawei grant HO2017050001B8s.

References:

[1] Nuamah, K. (2021). *Deep Algorithmic Question Answering: Towards a Compositionally Hybrid AI for Algorithmic Reasoning*. KR 2021 Workshop on Knowledge Representation for Hybrid and Compositional AI (KRHCAI).
[2] Bundy, A.; Nuamah, K. (2022). Unified Decomposition-Aggregation (UDA) Rules: Dynamic, Schematic, Novel Axioms. Conferences on Intelligent Computer Mathematics (CICM 2022).
[3] Brimble, G. (2021). *Automatic knowledge discovery*. UG Project Dissertation, School of Informatics, University of Edinburgh.
[4] Fletcher, T.; Bundy, A.; Nuamah, K. (2021). *GPy-ABCD: A Configurable Automatic Bayesian Covariance Discovery Implementation*. 8th ICML Workshop on Automated Machine Learning (AutoML).

School of informatics