Learning Contextualized Knowledge Structures for Commonsense Reasoning

Jun Yan, Mrigank Raman, Tianyu Zhang, Ryan Rossi, Handong Zhao, Sungchul Kim, Nedim Lipka, Xiang Ren

University of Southern California   **Indian Institute of Technology Delhi**
Tsinghua University   **Adobe Research**

Problem

❑ **KG-Augmented Commonsense Question Answering**

Input:

Question: Printing on a printer can get expensive because it does what?
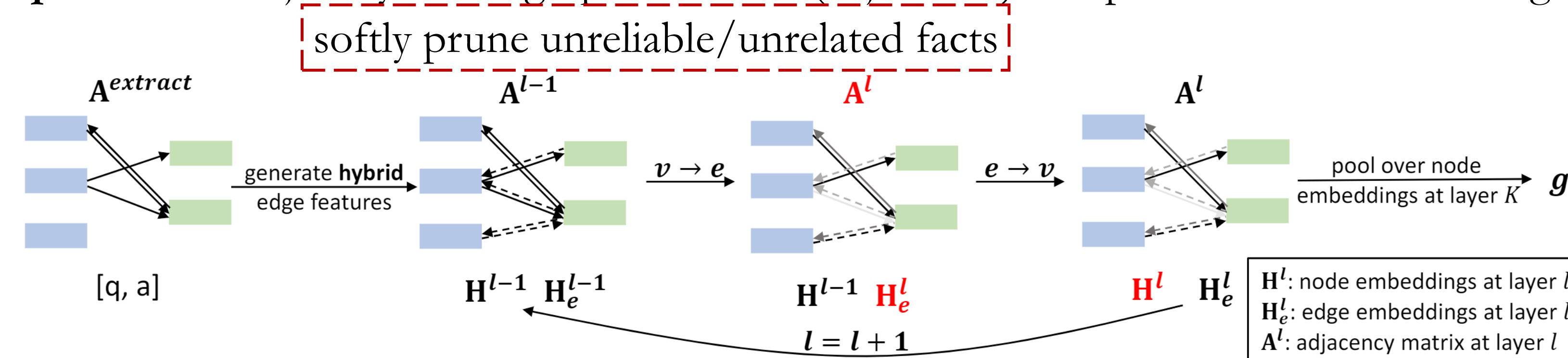Choices: A. explode B. use paper C. store information D. queue E. noise

+

ConceptNet
A commonsense knowledge graph

Output:

B. use paper

Background

❑ **Workflow**   Contextualized KG: G

Graph Initialization

$[q, a]$ → Text Encoder → s → concat → MLP → $\rho(q, a)$

plausibility score

❑ **Prior Work**

Graph Initialization
▪ Node Set: recognized concepts
▪ Edge Set: extracted edges (facts) from ConceptNet
Graph Encoder: operates on static graph structure.

❑ **Challenges**

▪ Missing facts
▪ Unreliable facts
▪ Unrelated facts

Hybrid Graph Network (HGN)

Graph Initialization
▪ Node Set: recognized concepts   complete missing facts
▪ Edge Set: extracted edges from ConceptNet + generated edges → **hybrid edge features**
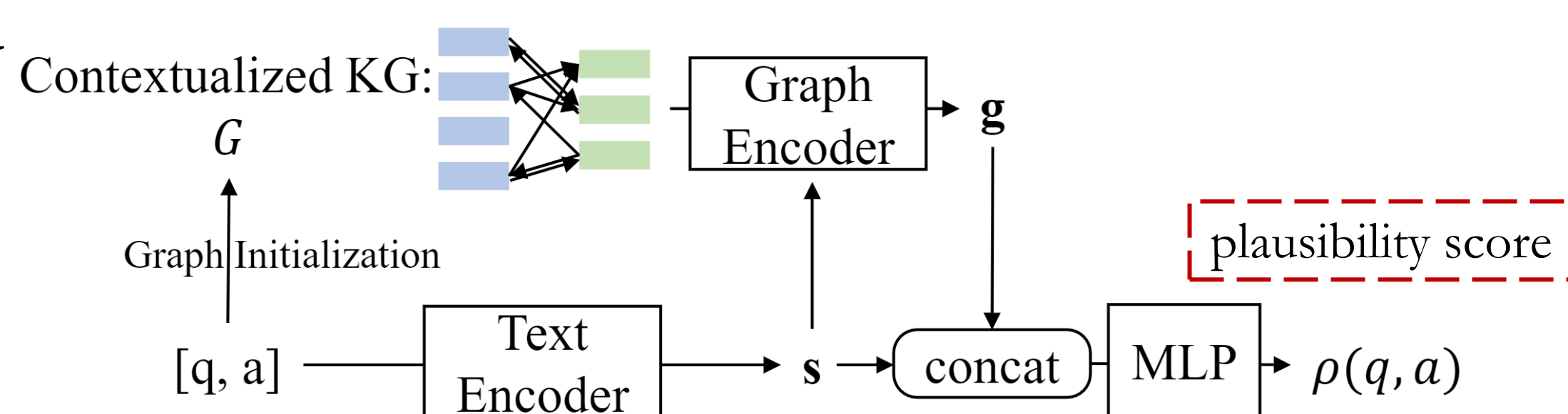**Graph Encoder:** jointly learns graph structure (adj matrix) and parameters for reasoning.

softly prune unreliable/unrelated facts

$\mathbf{A}^{extract}$ $\mathbf{A}^{l-1}$ $\mathbf{A}^l$ $\mathbf{A}^l$

generate **hybrid** edge features $v \to e$ $e \to v$ pool over node embeddings at layer K $\mathbf{g}$

$[q, a]$ $\mathbf{H}^{l-1}$ $\mathbf{H}_e^{l-1}$ $\mathbf{H}^{l-1}$ $\mathbf{H}_e^l$ $\mathbf{H}^l$ $\mathbf{H}_e^l$

$l = l + 1$

$\mathbf{H}^l$: node embeddings at layer l
$\mathbf{H}_e^l$: edge embeddings at layer l
$\mathbf{A}^l$: adjacency matrix at layer l

$$\mathcal{L} = \sum_{(q,a,y) \in D_{train}} \left[L_{task}(q, a, y) + \beta \cdot L_{prune}(\mathbf{A}^K) \right]$$ sparsity regularization

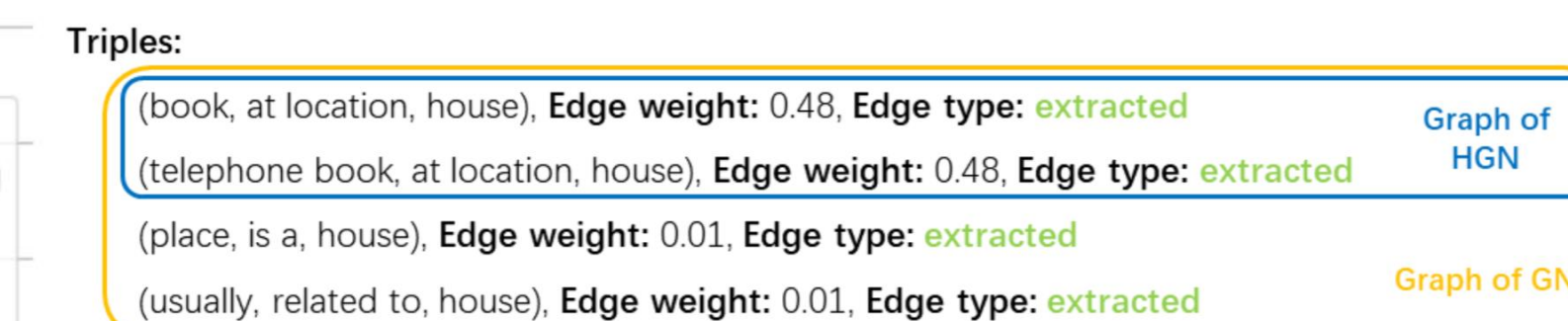## Experiments

❑ **Performance**

CommonsenseQA   OpenbookQA
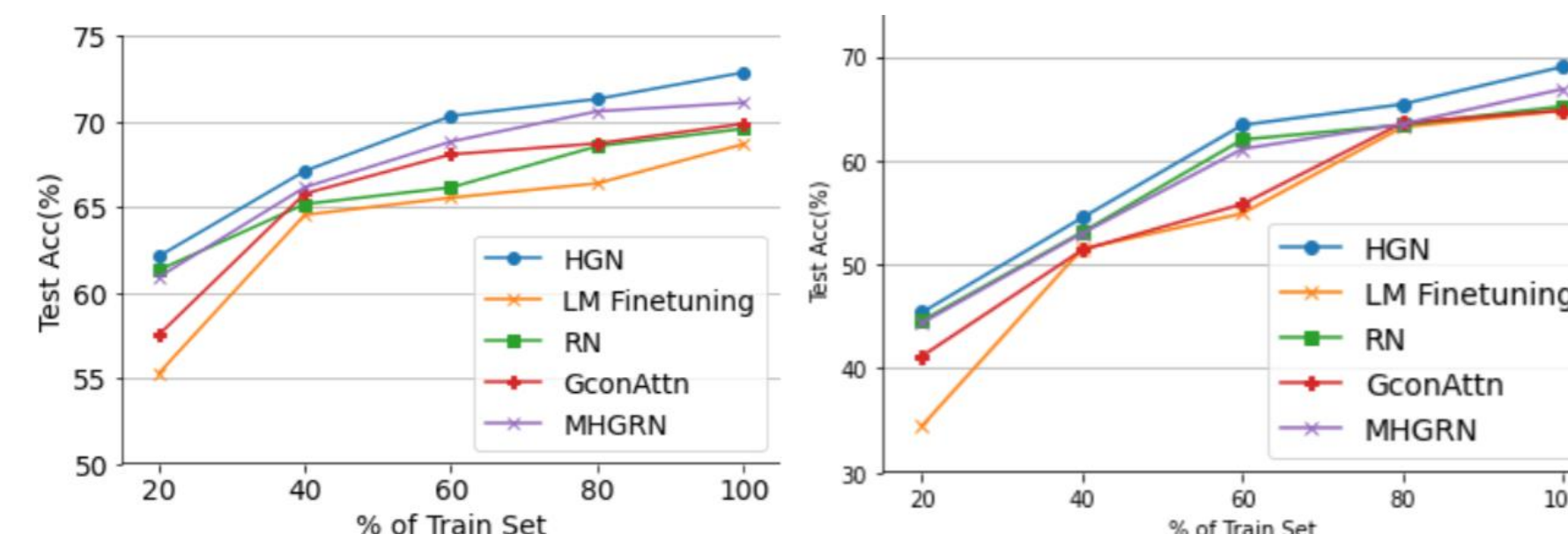
HGN achieves performance gain on commonsense reasoning benchmarks with superior contextualized knowledge structures

❑ **Case Studies**

Question: What is a place that usually does not have an elevator and that sometimes has a telephone book?

Answer: house

Triples:

(book, at location, house), **Edge weight:** 0.48, **Edge type:** extracted
(telephone book, at location, house), **Edge weight:** 0.48, **Edge type:** extracted
(place, is a, house), **Edge weight:** 0.01, **Edge type:** extracted
(usually, related to, house), **Edge weight:** 0.01, **Edge type:** extracted

Graph of HGN   Graph of GN

Question: Where would you find an office worker gossiping with their colleagues?

Answer: water cooler

Triples:

(gossip, related to, water cooler), **Edge weight:** 0.09, **Edge type:** extracted
(office, related to, cooler), **Edge weight:** 0.09, **Edge type:** extracted
(office, related to, water), **Edge weight:** 0.09, **Edge type:** extracted
(office, related to, water cooler), **Edge weight:** 0.09, **Edge type:** extracted
(office worker, is located at, water cooler), **Edge weight:** 0.02, **Edge type:** generated
(worker, is located at, water cooler), **Edge weight:** 0.02, **Edge type:** generated
(gossiping, is located at, water cooler), **Edge weight:** 0.02, **Edge type:** generated

Graph of GN   Graph of HGN