



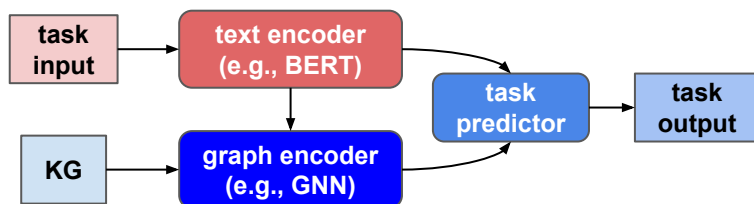
SalKG: Learning From Knowledge Graph Explanations for Commonsense Reasoning

ICML

Aaron Chan¹, Boyuan Long^{*1}, Jiashu Xu^{*1}, Soumya Sanyal¹, Tanishq Gupta², Xiang Ren¹

¹University of Southern California, ²IIT Delhi (*equal contribution)

1. What is a KG-augmented model?



Advantages of KG-augmented models:

- Perform better on knowledge-intensive tasks.
- Can “explain” their predictions via attention over KG.

2. Background

Limitations of KG-Augmented Models

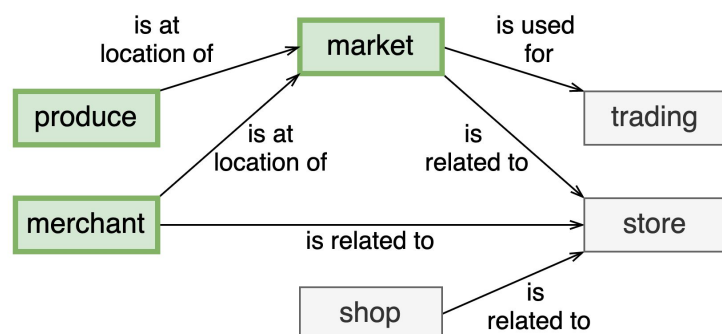
- KG may not be useful for given task instance.
- Even if KG is useful, not all parts of the KG may be.
- Existing KG-augmented models are not explicitly taught when KG or specific parts of KG should be used.

KG Saliency Explanations

- For each instance, indicate which KG inputs are most useful (i.e., push model to make the correct prediction).
- Created via *saliency methods* (Bastings, 2019).

What kind of store does a merchant have if they sell produce?

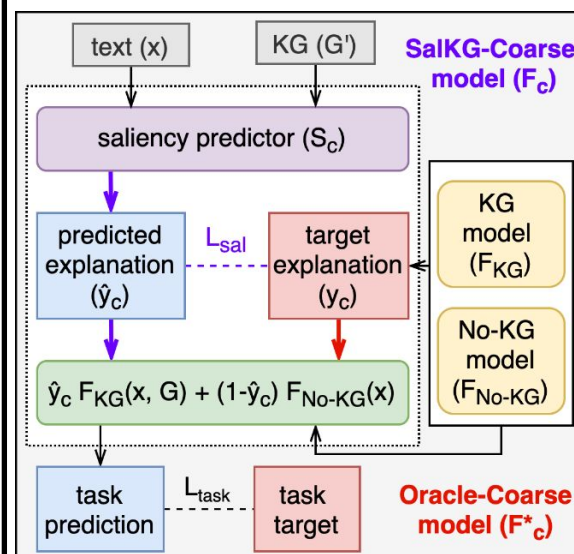
A. the sun B. market C. business D. mall E. shopping center



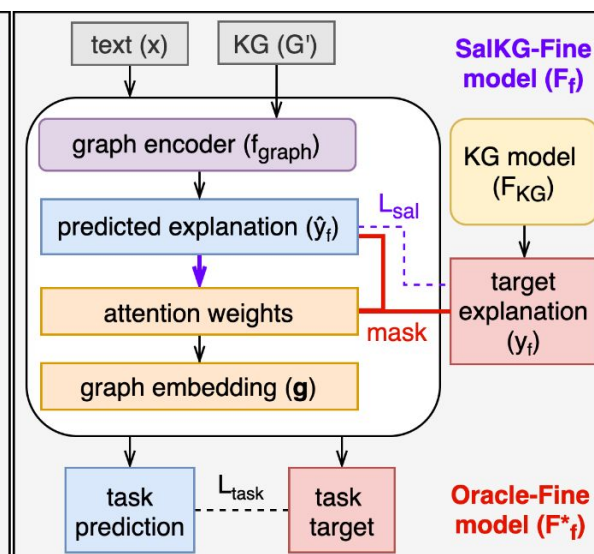
KG nodes `market`, `produce`, and `merchant` most strongly influence the model to predict the correct answer `market`.

3. SalKG

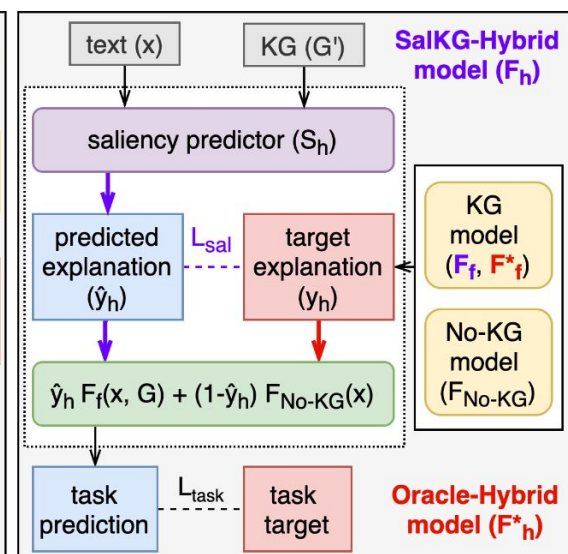
Step 1: Train vanilla KG-augmented (KG) and non-KG-augmented (No-KG) models on the given task.



Step 2: Create **coarse** (Is the KG useful?) and **fine** (Which nodes/paths in the KG are useful?) saliency explanations from KG and No-KG models, using saliency methods.

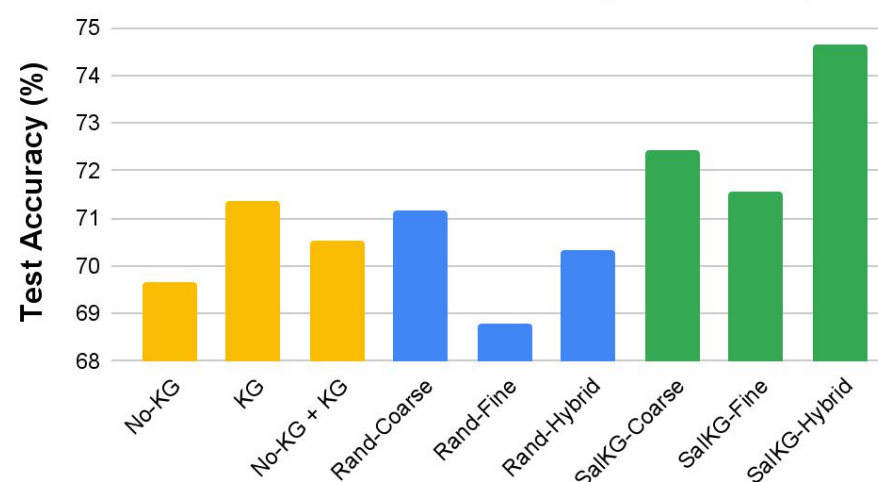


Step 3: Jointly train KG model to predict coarse and/or fine explanations, then solve task by attending to KG features highlighted by predicted explanations.



4. Results

Performance on CommonsenseQA (RoBERTa+RN)



- **Hybrid** = uses both coarse and fine explanations
- **No-KG + KG** = simple mean ensemble of No-KG and KG
- **Rand** = SalKG variant that uses random explanations

5. Conclusion

Takeaways

- Saliency explanations can provide strong signal for teaching KG-augmented models to focus on useful KG features.
- Coarse and fine explanations are *complementary*.

Future Work

- Incorporate active learning into SalKG, so models can also get human feedback about KG saliency.
- Design new ways to train model via KG explanations.

6. Resources

Paper: arxiv.org/abs/2104.08793

Contact: Aaron Chan (chanaaro@usc.edu)

