Towards Factual Language Models: Knowledge Graph Integration with Pre-Trained Language Models

Alaa Almutawa (alaaj.almutawa@gmail.com)
Koumudi Ganepola Achchige (koumudiganepola@gmail.com)
Barcelona, Spain - eBISS 2023

1. Motivation

PLMs have proved to be particularly useful in general natural language processing tasks, they have also demonstrated inconsistencies in responses, factual errors and a lack of contextual awareness

Why Knowledge Graph enhanced PLM ? Leverage the wealth of structured data contained within Knowledge Graphs (KG) to address the consistency, factual awareness as well as contextual-awareness issues that PLMs demonstrate.

What happens when you ask ChatGPT
- Who wrote System R paper ?
- Who authored System R research paper ?

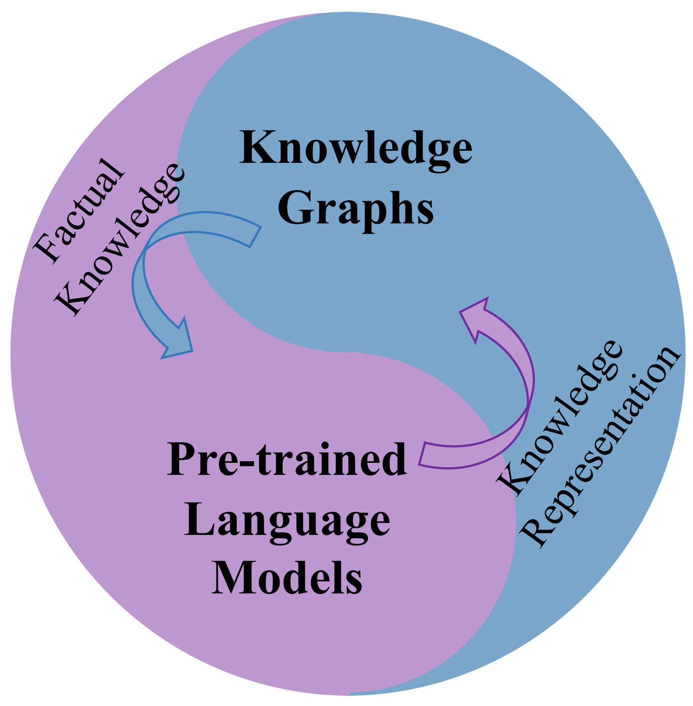

Figure 1: KGs are factual, structured and consistent. While PLMs are exceptional for NLP tasks. Synergizing KG and PLM can provide a path to improving each of the technologies. [1]

2. Knowledge Injection in PLM

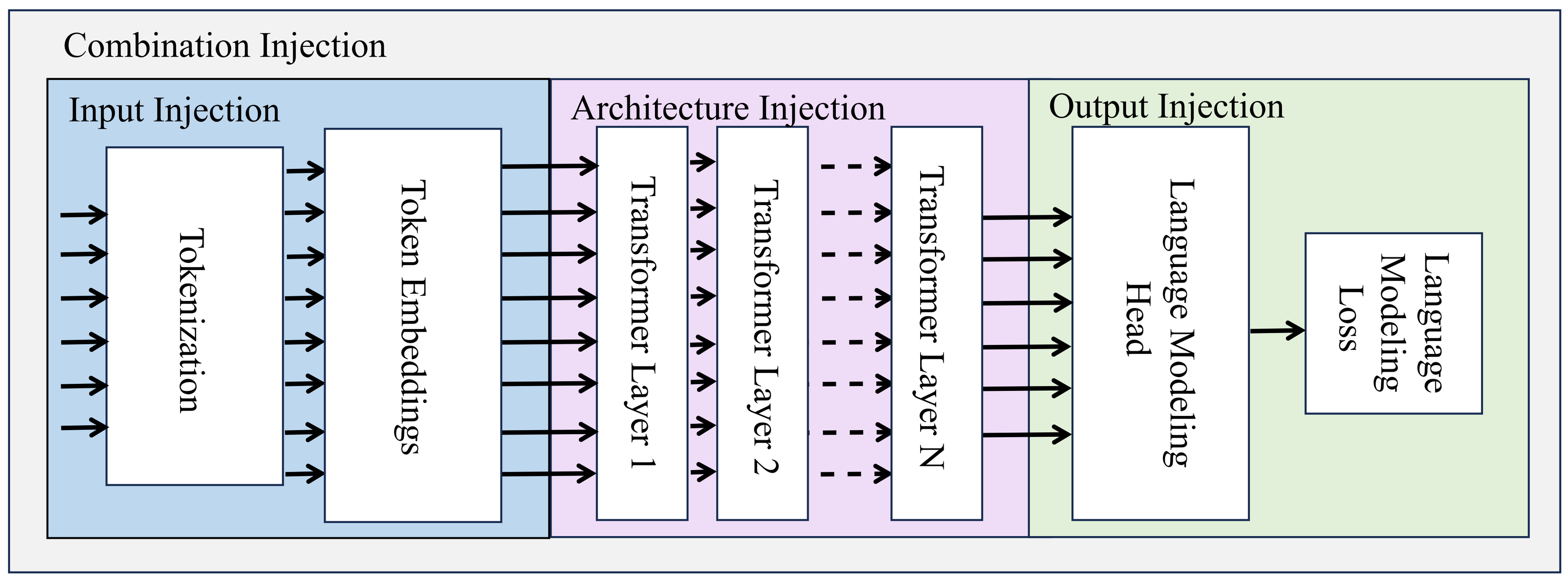

Figure 2: Categories of knowledge injection in PLM. [2]

1. **Input Injection:** methods that modify the input of the PLM.
   Ex: Alignment of text to KG triplets.
2. **Architecture Injection:** methods that modify PLM architecture through introduction of new layers, or modification of the existing layers.
   Ex: Modification of attention mechanism.
3. **Output Injection:** methods that modify the output or the loss that is used in the PLM.

3. Overview

Models	Input	Modifications to Architecture	New Pre-training Tasks	Trainable Parameters	Base PLM	Base PLM Parameters
ERNIE (Jun 2019)	Entity Embeddings (TransE)	K-Encoder (Aggregator)	Entity Typing Relation Classification	114M	$\text{BERT}_{\text{BASE}}$	110M
K-BERT (April 2020)	K-Query	Knowledge Layer, Seeing Layer, Masked-Self-Attention	-	110M	$\text{BERT}_{\text{BASE}}$	110M
KEPLER (Nov 2020)	Entity Description	-	-	123M	$\text{RoBERTa}_{\text{BASE}}$	123M
K-Adapter (Dec 2020)	Entity Aligned Text	Factual Adapter	Relation Classificaion	42M	$\text{RoBERTa}_{\text{LARGE}}$	335M

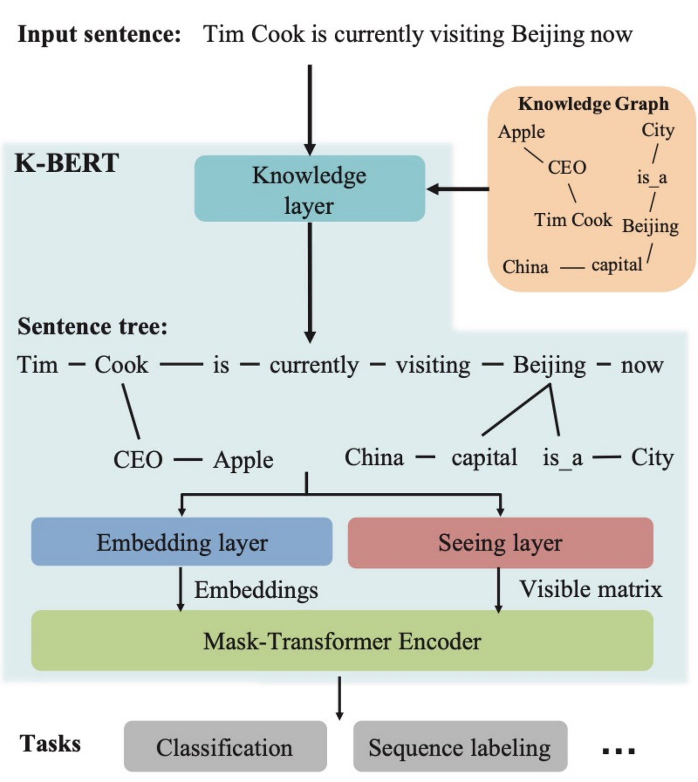

Figure 3: K-BERT [3]

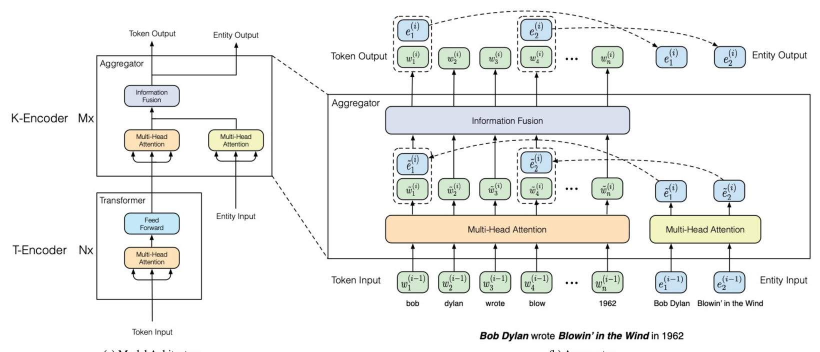

Figure 4: ERNIE [4]

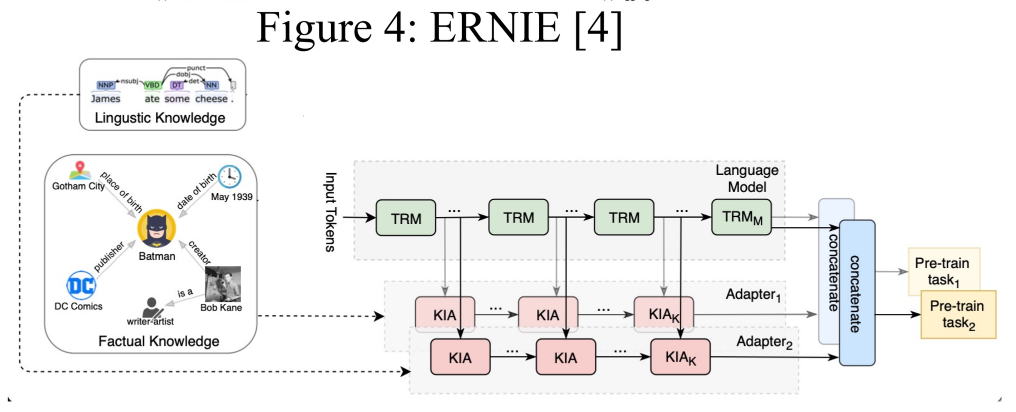

Figure 5: K-Adapter [5]

4. Comparison

Robustness: *Model's ability to remain consistent with the introduction of new triplets (subgraphs) in the KG.*
- K-BERT uses K-Query to query the KG and later injects the triplets to the input text creating a sentence tree.

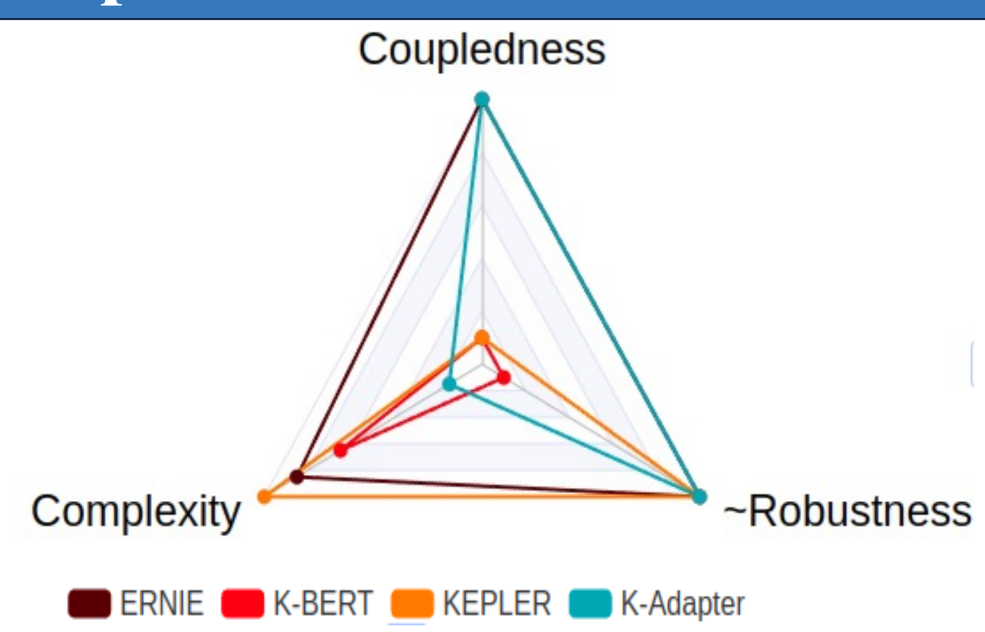

Figure 6: Dimensions to evaluate KG-enhanced PLMs.

Complexity: *Measured using the number of trainable parameters in the model*
- K-Adapter requires to train less number of parameters. Hence to integrate the PLM with KG, it requires less computational resources in comparasion to the other models.

Coupledness: *Model's ability to extrapolate unseen entities in the KG*
- ERNIE and K-Adapter introduce pre-training tasks demonstrating a high level of coupledness.
- Both K-BERT and KEPLER introduces KG only within the finetuning phase, they do not introduce any separate pre-training tasks.

5. Challenges

1. The **lack of a common benchmark** to evaluate the factual awareness
   ⟶ Inability to compare performance using quantitative measures.
2. Each model uses **different** KGs, evaluation tasks and text datasets
   ⟶ Difficult to compare these models even using their base PLM
3. **Computational performances not reported** (*i.e. runtime, resource usage)*
   ⟶ The complexity dimension is based on the number of trainable parameters rather than considering a more holistic view that includes space and time complexities.

6. Conclusion

- The **increasing popularity** of PLM usage in many applications, creates a need to ensure that PLMs are consistent and factual. KG integration offers new horizons to improve PLMs effectiveness.
- **Various** innovative methods have been explored by researchers. However, major challenges remain to fully synergize PLMs and KGs.
- We discuss different **KG-enhanced PLMs**: ERNIE, K-BERT, KEPLER and K-Adapter. Then compare each against defined dimensions: Coupledness, Complexity, Roubustness.
- **K-BERT** outperforms all other models in coupledness (low) and robustness (high) dimensions while having an intermediate-level complexity.

7. References

[1] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J. & Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. (2023)
[2] Colon-Hernandez, P., Havasi, C., Alonso, J., Huggins, M., & Breazeal, C. Combining pre-trained language models and structured knowledge. arXiv preprint arXiv:2101.12294. (2021)
[3] W. Liu et al., "K-BERT: Enabling language representation with knowledge graph," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 03, pp. 2901–2908, 2020
[4] Z. Zhang et al., "ERNIE: Enhanced language representation with informative entities," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019
[5] R. Wang et al., "K-Adapter: Infusing knowledge into pre-trained models with adapters," 2020
[6] X. Wang et al., "KEPLER: A unified model for knowledge embedding and pre-trained language representation," Transactions of the Association for Computational Linguistics 9, 2021