QA-GNN: Question Answering using Language Models and Knowledge Graphs

AILAB 백형렬

QAGNN

해당 논문 선정이유
- 해당 논문은 Knowledge graph를 활용하여 QA task에 적용
- 텍스트 데이터에 그래프 모델을 QA에 적용 방법 리서치(특히 Node, Edge 구성)

해당 논문 핵심
- Graph node, edge
- Relevance scoring
- Reasoning

QAGNN – Node, Edge

- Question (Q), Answer (A) text로 Context Node생성
- Q, A의 Entity로 Node 생성
- Q <-> A Edge 연결
- Q <-> A 사이에 k-hop Other Node 생성
- a) Edge type (Node 간 relation), b) Other Node 등은 ConceptNet 사용

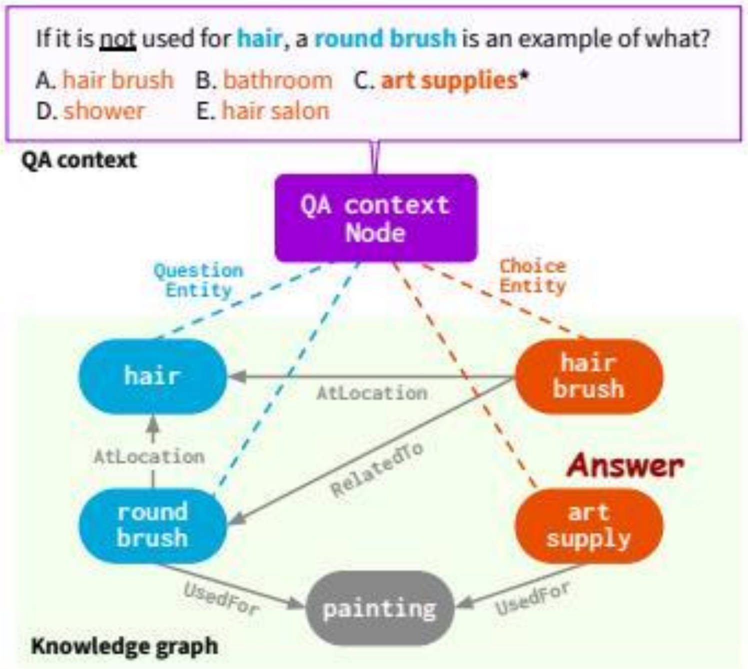


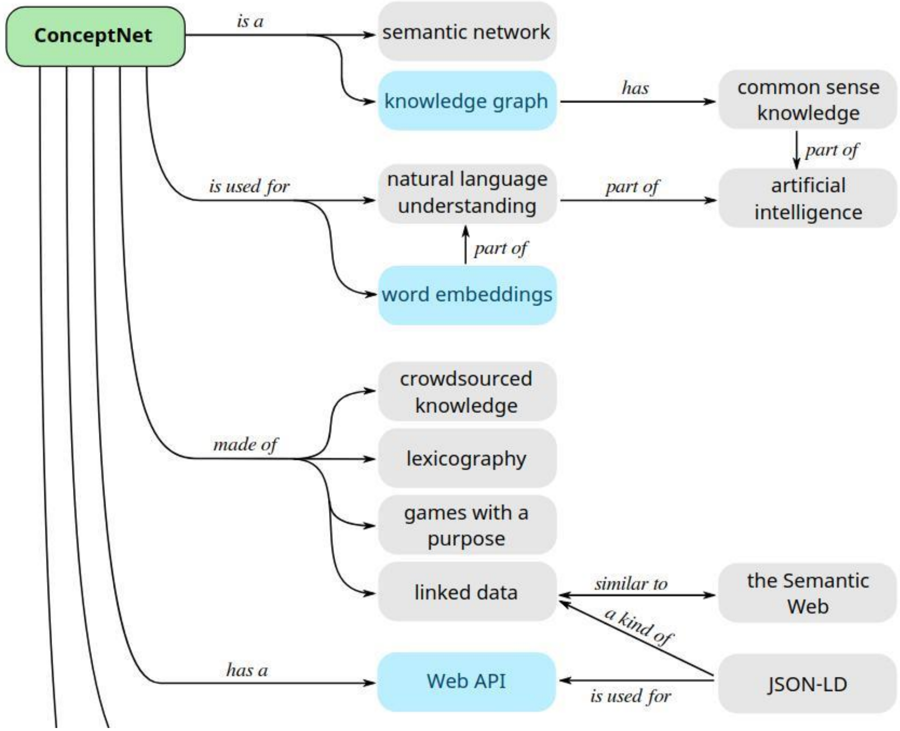
Figure 1: Given the QA context (question and answer choice; purple box), we aim to derive the answer by performing joint reasoning over the language and the knowledge graph (green box).

QAGNN – ConceptNet

- Knowledge graph. 2004년 처음 구축[1]. 현재 ver 5.5까지 공개[2].
- triplet (start node, relation, end node) 을 그래프로 연결
- 21M edges, 8M nodes, Multilingual vocab

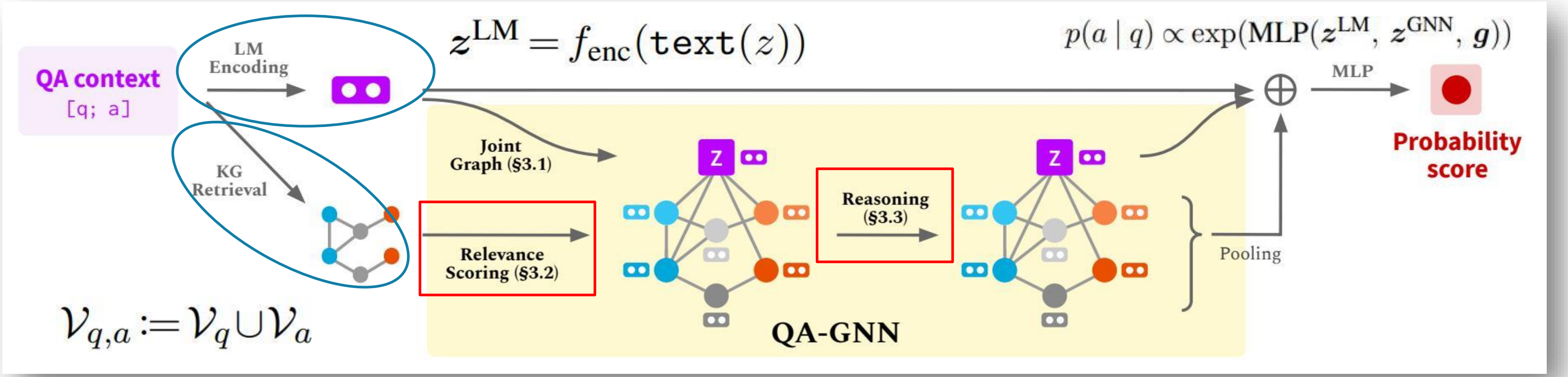
The core relations are:

- **Symmetric relations**: *Antonym*, *DistinctFrom*, *EtymologicallyRelatedTo*, *LocatedNear*, *RelatedTo*, *SimilarTo*, and *Synonym*
- **Asymmetric relations**: *AtLocation*, *CapableOf*, *Causes*, *CausesDesire*, *CreatedBy*, *DefinedAs*, *DerivedFrom*, *Desires*, *Entails*, *ExternalURL*, *FormOf*, *HasA*, *HasContext*, *HasFirstSubevent*, *HasLastSubevent*, *HasPrerequisite*, *HasProperty*, *InstanceOf*, *IsA*, *MadeOf*, *MannerOf*, *MotivatedByGoal*, *ObstructedBy*, *PartOf*, *ReceivesAction*, *SenseOf*, *SymbolOf*, and *UsedFor*



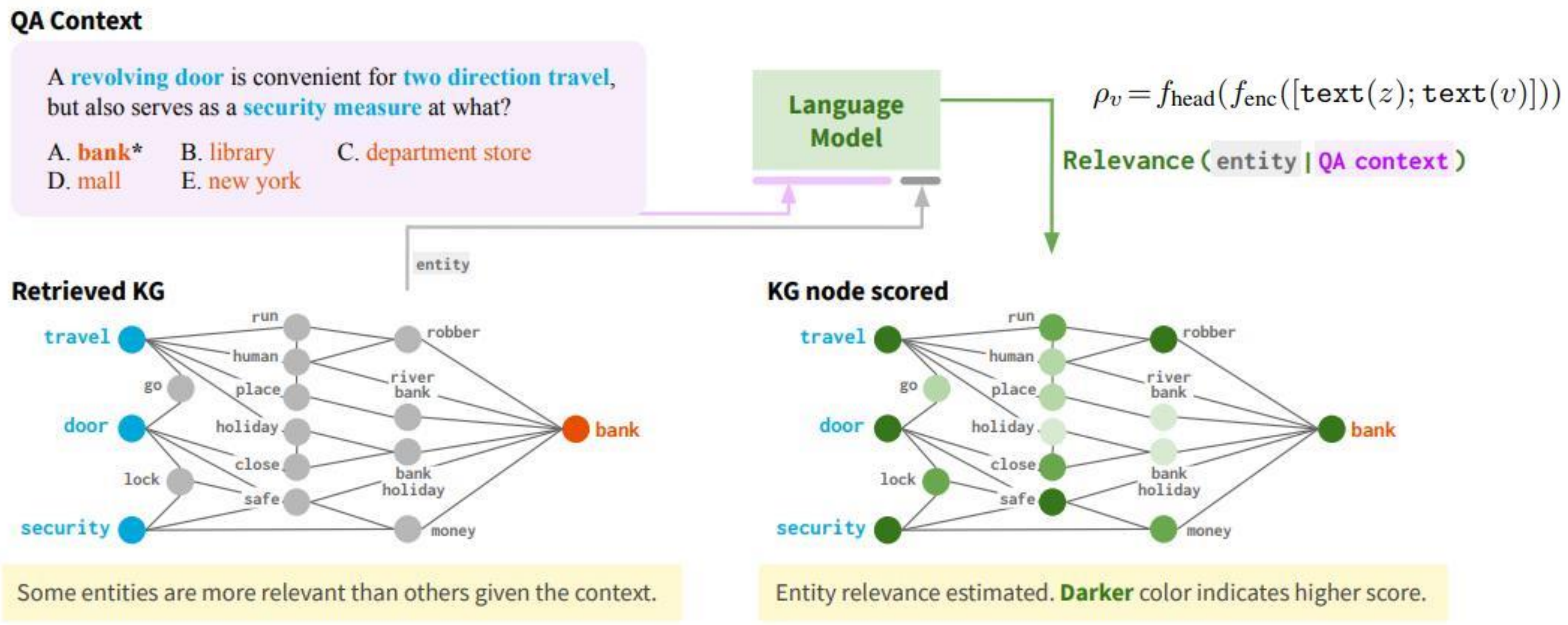
1) Liu, H., and Singh, P. 2004. ConceptNet – a practical commonsense reasoning tool-kit. BT Technology Journal 22(4):211–226.
2) Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In AAAI Conference on Artificial Intelligence, pages 4444–4451.

QAGNN - Method

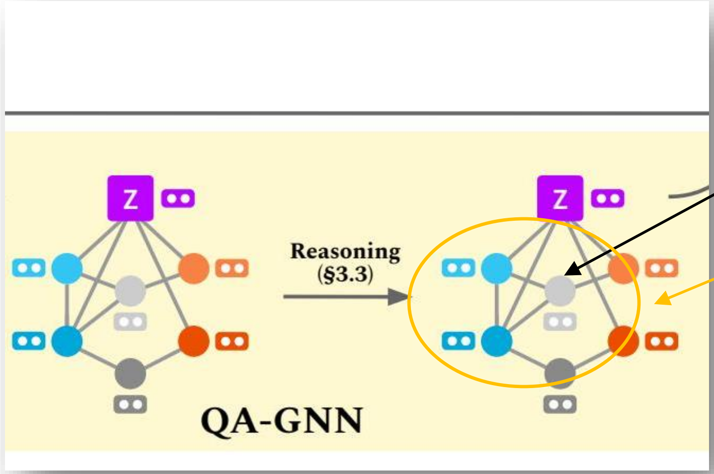


- Conext node와 Q, A의 Entity node 연결
- Node representation update시, node 간 relevance 고려 (Relevance Scoring)
- Node representation update하면서 reasoning

QAGNN – Method: Relevance scoring



QAGNN – Method: Reasoning



$$h_t^{(\ell+1)} = f_n\left(\sum_{s\in\mathcal{N}_t\cup\{t\}} \alpha_{st} m_{st}\right) + h_t^{(\ell)}$$

s: neighbor nodes
u_s: node s의 type (e.g. context, Q, A, Others) Emb.
r_st: node s, t edge rel. Emb.

$$m_{st} = f_m(h_s^{(\ell)}, u_s, r_{st})$$

$$\alpha_{st} = \frac{\exp(\gamma_{st})}{\sum_{t'\in\mathcal{N}_s\cup\{s\}}\exp(\gamma_{st'})}, \quad \gamma_{st} = \frac{q_s^\top k_t}{\sqrt{D}}$$

$$q_s = f_q(h_s^{(\ell)}, u_s, \rho_s),$$
$$k_t = f_k(h_t^{(\ell)}, u_t, \rho_t, r_{st})$$

- Neighbor nodes의 rep., type, relation edge, relevance score 고려하여 node t 업데이트
- Node t representation: h_t
- Message m_st: node t의, neighbor node s의 msg representation
    - msg rep.: node rep., node type, edge relation으로 계산
- Node representation update시, node 간 relevance 고려 (Relevance Scoring)

QAGNN – Experiments

Methods	Test
RoBERTa (Liu et al., 2019)	72.1
RoBERTa+FreeLB (Zhu et al., 2020) (ensemble)	73.1
RoBERTa+HyKAS (Ma et al., 2019)	73.2
RoBERTa+KE (ensemble)	73.3
RoBERTa+KEDGN (ensemble)	74.4
XLNet+GraphReason (Lv et al., 2020)	75.3
RoBERTa+MHGRN (Feng et al., 2020)	75.4
Albert+PG (Wang et al., 2020b)	75.6
Albert (Lan et al., 2020) (ensemble)	76.5
UnifiedQA* (Khashabi et al., 2020)	**79.1**
RoBERTa + QA-GNN (**Ours**)	76.1

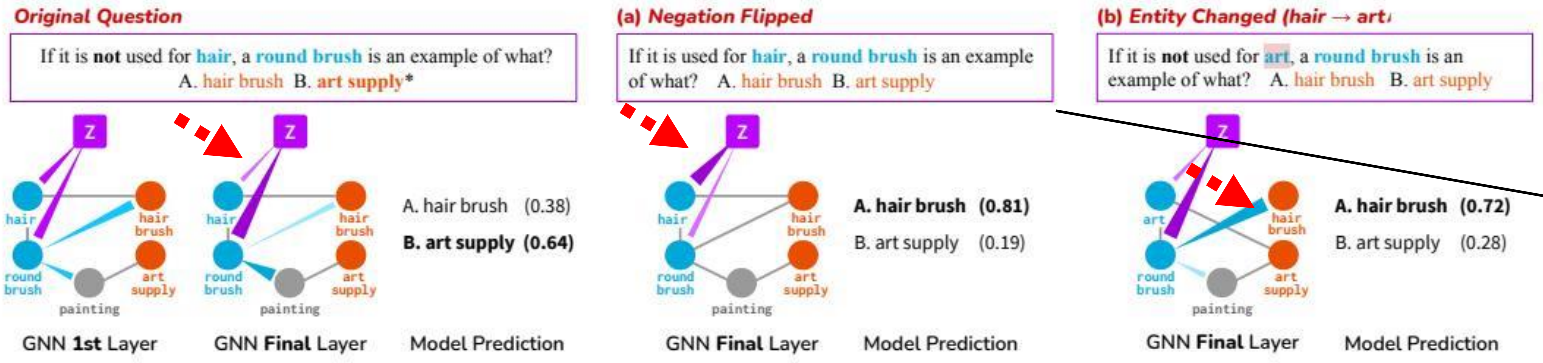
Table 3: **Test accuracy on *CommonsenseQA*'s official leaderboard.** The top system, UnifiedQA (11B parameters) is 30x larger than our model.

Methods	Test
Careful Selection (Banerjee et al., 2019)	72.0
AristoRoBERTa	77.8
KF + SIR (Banerjee and Baral, 2020)	80.0
AristoRoBERTa + PG (Wang et al., 2020b)	80.2
AristoRoBERTa + MHGRN (Feng et al., 2020)	80.6
Albert + KB	81.0
T5* (Raffel et al., 2020)	83.2
UnifiedQA* (Khashabi et al., 2020)	**87.2**
AristoRoBERTa + QA-GNN (**Ours**)	82.8

Table 5: **Test accuracy on *OpenBookQA* leaderboard.** All listed methods use the provided science facts as an additional input to the language context. The top 2 systems, UnifiedQA (11B params) and T5 (3B params) are 30x and 8x larger than our model.

- Dataset: OpenBookQA, Commonsense QA
- parameter 수 고려 좋은 성적

QAGNN – Experiments: Structured reasoning



Negation, Substitution에 강건

Figure 5: **Analysis of QA-GNN's behavior for structured reasoning**. Given an original question (left), we modify its negation (middle) or topic entity (right): we find that QA-GNN adapts attention weights and final predictions accordingly, suggesting its capability to handle structured reasoning.

Example (Original taken from *CommonsenseQA* Dev)	RoBERTa Prediction	Our Prediction
[Original] If it is **not** used for hair, a round brush is an example of what? A. hair brush B. art supply	A. hair brush (✗)	B. art supply (✓)
[Negation flip] If it is used for hair, a round brush is an example of what?	A. hair brush (✓ just no change?)	A. hair brush (✓)
[Entity change] If it is not used for art a round brush is an example of what?	A. hair brush (✓ just no change?)	A. hair brush (✓)
[Original] If you have to read a book that is very dry you may become what? A. interested B. bored	B. bored (✓)	B. bored (✓)
[Negation ver 1] If you have to read a book that is very dry you may **not** become what?	B. bored (✗)	A. interested (✓)
[Negation ver 2] If you have to read a book that is **not** dry you may become what?	B. bored (✗)	A. interested (✓)
[Double negation] If you have to read a book that is **not** dry you may **not** become what?	B. bored (✓ just no change?)	A. interested (✗)

Table 8: **Case study of structured reasoning**, comparing predictions by RoBERTa and our model (RoBERTa + QA-GNN). Our model correctly handles changes in negation and topic entities.

Conclusion

- QA에 KB(conceptnet) joint
- entity와 QA context의 relevance, entity간 relation, node type 사용하여 QA 해결
- 적은 parameter로 좋은 성적 달성