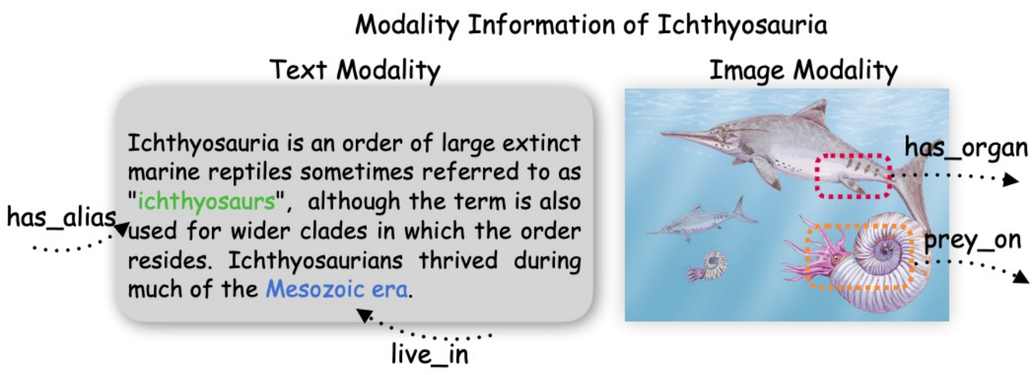
# Multiple Heads Are Better Than One: Mixture of Modality Knowledge Experts for Entity Representation Learning

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu,
Wen Zhang, Huajun Chen
**Zhejiang University & Ant Group**




Case Figure: Different modalities can represent various aspects of entity information, and information within the same modality can also play different roles depending on the relational context

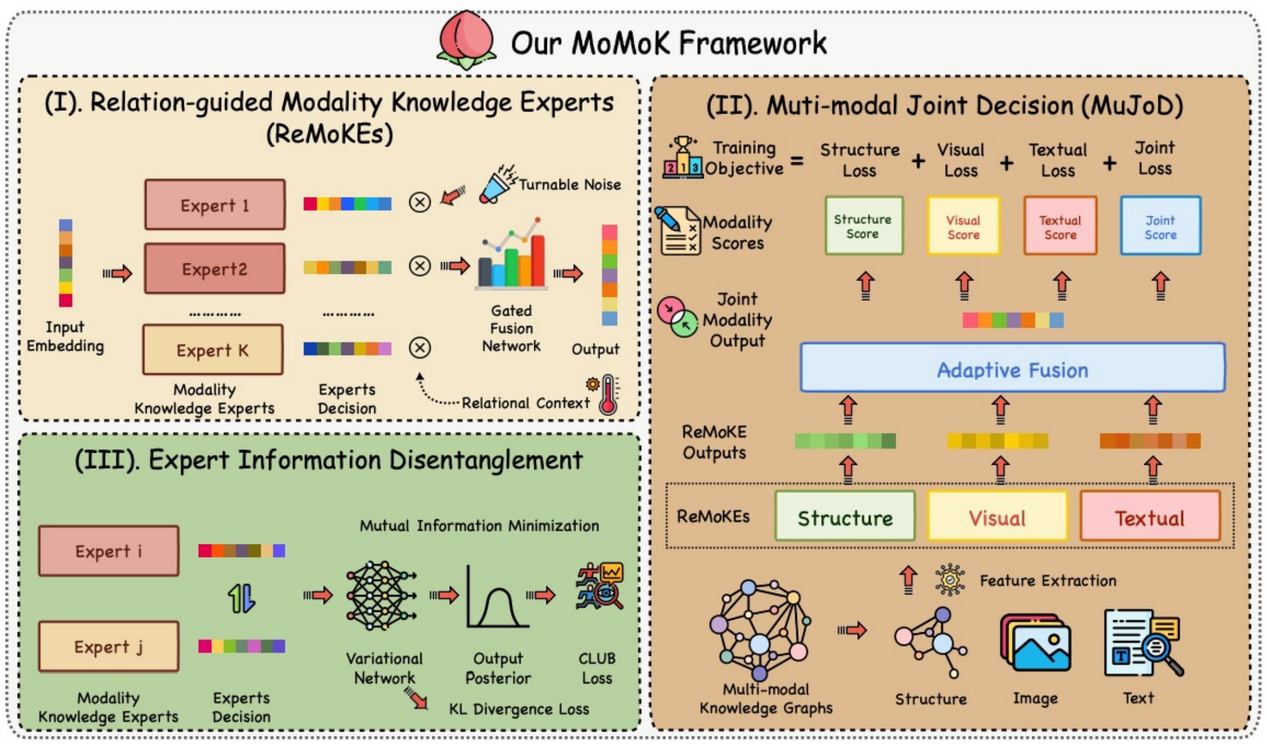
**Multi-modal Knowledge Graph Completion (MMKGC)**
**Existing MMKGC Paradigms:**

1. Entity encoding with BERT / VGG / ViT
2. Multi-modal fusion in the same representation space
3. Triple scoring with traditional KGE / KGC models like TransE

**Challenges**:
Overlooks the information **diversity** both **inter-modality and intra-modality**.

**What we NEED?**
A more fine-grained MMKGC framework that modality functioning as a senior expert, gathering the insights of junior experts within the corresponding modality.

**Our Framework**: **M**ixture **o**f **Mo**dality **K**nowledge experts **(MOMOK)**




**Module1: Relation-Guided Modality Knowledge Experts**

- Adaptive Multi-modal Fusion with Relational Contexts and Mixture-of-Experts Module

$$G_i(\mathcal{V}^e_{m,i}, r) = \frac{\exp\left((\mathcal{U}_m(\mathcal{V}^e_{m,i}) + \delta_{m,i})/\sigma(\varepsilon_r)\right)}{\sum_{j=1}^{K}\exp\left((\mathcal{U}_m(\mathcal{V}^e_{m,j}) + \delta_{m,j})/\sigma(\varepsilon_r)\right)},$$

**Module2: Multi-modal Joint Decision**

- Achieve Inter-modal and Intra-modal Joint Decision

$$\widehat{\boldsymbol{e}}_{Joint,r} = \frac{\exp(\mathcal{W}_{attn} \odot \mathcal{P}_m(\widehat{\boldsymbol{e}}_{m,r}))}{\sum_{n\in\mathcal{M}}\exp(\mathcal{W}_{attn} \odot \mathcal{P}_n(\widehat{\boldsymbol{e}}_{n,r}))}\mathcal{P}_m(\widehat{\boldsymbol{e}}_{m,r})$$

**Module3: Expert Information Disentanglement**

- Mutual Information Estimation to Guide the Experts Learning
- Make the Experts more Diverse and Functional

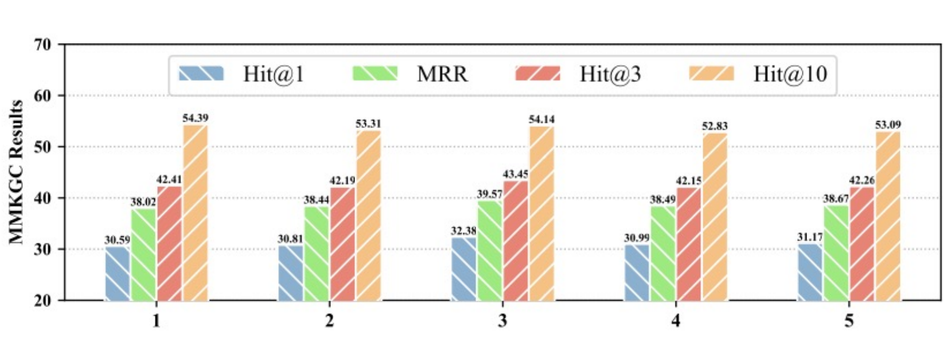
$$\mathcal{L}_{club} = \frac{1}{K^2}\sum_{m\in\mathcal{M}}\sum_{e\in\mathcal{B}}\sum_{i=1}^{K}\sum_{j\neq i}^{K}\left(\log\mathcal{Q}_{\theta,m}(\mathcal{V}^e_{m,j}|\mathcal{V}^e_{m,i}) - \sum_{e'\in\mathcal{B}-\{e\}}\log\mathcal{Q}_{\theta,m}(\mathcal{V}^{e'}_{m,j}|\mathcal{V}^e_{m,i})\right)$$

## Experiments and Evaluation

| Model | MKG-W MRR | MKG-W Hit@1 | MKG-Y MRR | MKG-Y Hit@1 | DB15K MRR | DB15K Hit@1 | DB15K Hit@3 | DB15K Hit@10 | KVC16K MRR | KVC16K Hit@1 | KVC16K Hit@3 | KVC16K Hit@10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Uni-modal KGC Methods* | | | | | | | | | | | | |
| **TransE** | 29.19 | 21.06 | 30.73 | 23.45 | 24.86 | 12.78 | 31.48 | 47.07 | 8.54 | 0.64 | 10.97 | 23.42 |
| **DistMult** | 20.99 | 15.93 | 25.04 | 19.33 | 23.03 | 14.78 | 26.28 | 39.59 | 6.37 | 3.03 | 6.11 | 12.61 |
| **ComplEx** | 24.93 | 19.09 | 28.71 | 22.26 | 27.48 | 18.37 | 31.57 | 45.37 | 12.85 | 7.48 | 13.79 | 23.18 |
| **RotatE** | 33.67 | 26.80 | 34.95 | 29.10 | 29.28 | 17.87 | 36.12 | 49.66 | 14.33 | 8.25 | 15.37 | 26.17 |
| **PairRE** | 34.40 | 28.24 | 32.01 | 25.53 | 31.13 | 21.62 | 35.91 | 49.30 | - | - | - | - |
| **TuckER** | 29.59 | 23.93 | 37.05 | 34.59 | 33.86 | 25.34 | 37.91 | 50.38 | 15.90 | 9.79 | 17.24 | 27.58 |
| *Multi-modal KGC Methods* | | | | | | | | | | | | |
| **IKRL** | 32.36 | 26.11 | 33.22 | 30.37 | 26.82 | 14.09 | 34.93 | 49.09 | 11.11 | 5.42 | 11.46 | 22.39 |
| **TBKGC** | 31.48 | 25.31 | 33.99 | 30.47 | 28.40 | 15.61 | 37.03 | 49.86 | 5.39 | 0.35 | 5.04 | 15.52 |
| **TransAE** | 30.00 | 21.23 | 28.10 | 25.31 | 28.09 | 21.25 | 31.17 | 41.17 | 10.81 | 5.31 | 11.34 | 21.89 |
| **MMKRL** | 30.10 | 22.16 | 36.81 | 31.66 | 26.81 | 13.85 | 35.07 | 49.39 | 8.78 | 3.89 | 8.99 | 18.34 |
| **RSME** | 29.23 | 23.36 | 34.44 | 31.78 | 29.76 | 24.15 | 32.12 | 40.29 | 12.31 | 7.14 | 13.21 | 22.05 |
| **VBKGC** | 30.61 | 24.91 | 37.04 | 33.76 | 30.61 | 19.75 | 37.18 | 49.44 | 14.66 | 8.28 | 15.81 | 27.04 |
| **OTKGE** | 34.36 | 28.85 | 35.51 | 31.97 | 23.86 | 18.45 | 25.89 | 34.23 | 8.77 | 5.01 | 9.31 | 15.55 |
| **MoSE*** | 33.34 | 27.78 | 36.28 | 33.64 | 28.38 | 21.56 | 30.91 | 41.67 | 8.81 | 4.75 | 9.46 | 16.40 |
| **IMF*** | 34.50 | 28.77 | 35.79 | 32.95 | 32.25 | 24.20 | 36.00 | 48.19 | 12.01 | 7.42 | 12.82 | 21.01 |
| **QEB** | 32.38 | 25.47 | 34.37 | 29.49 | 28.18 | 14.82 | 36.67 | 51.55 | 12.06 | 5.57 | 13.03 | 25.01 |
| **VISTA** | 32.91 | 26.12 | 30.45 | 24.87 | 30.42 | 22.49 | 33.56 | 45.94 | 11.89 | 6.97 | 12.66 | 21.27 |
| **AdaMF** | 34.27 | 27.21 | **38.06** | 33.49 | 32.51 | 21.31 | 39.67 | 51.68 | 15.26 | 8.56 | 16.71 | 28.29 |
| *Negative Sampling Methods* | | | | | | | | | | | | |
| **MANS** | 30.88 | 24.89 | 29.03 | 25.25 | 28.82 | 16.87 | 36.58 | 49.26 | 10.42 | 5.21 | 11.01 | 20.45 |
| **MMRNS** | 35.03 | 28.59 | 35.93 | 30.53 | 32.68 | 23.01 | 37.86 | 51.01 | 13.31 | 7.51 | 14.19 | 24.68 |
| **MoMoK** | **35.89** | **30.38** | 37.91 | **35.09** | **39.57** | **32.38** | **43.45** | **54.14** | **16.87** | **10.53** | **18.26** | **29.20** |
| Improvements | +2.5% | +4.2% | - | +3.9% | +21.1% | +33.8% | +9.5% | +4.8% | +10.6% | +23.0% | +9.3% | +3.21% |

Main MMKGC results on 4 datasets

| Setting | | MKG-W MRR | MKG-W Hit@1 | DB15K MRR | DB15K Hit@1 | DB15K Hit@3 | DB15K Hit@10 |
|---|---|---|---|---|---|---|---|
| Full Model | | 35.89 | 30.38 | 39.57 | 32.38 | 43.45 | 54.14 |
| Modality Contribution | (1.1). Structure Modality | 32.82 | 27.73 | 36.45 | 29.36 | 39.99 | 49.86 |
| | (1.2). Image Modality | 32.75 | 27.78 | 36.84 | 29.80 | 40.10 | 50.42 |
| | (1.3). Text Modality | 32.62 | 27.66 | 37.04 | 29.93 | 40.49 | 50.39 |
| | (1.4). Joint Modality | 34.76 | 29.33 | 36.87 | 29.90 | 42.44 | 53.93 |
| Model Design | (2.1). w/o relational $\epsilon_r$ | 35.50 | 29.98 | 39.40 | 31.47 | 43.19 | 52.88 |
| | (2.2). w/o noise $\delta_m$ | 35.31 | 29.69 | 39.43 | 31.54 | 43.32 | 53.75 |
| | (2.3). w/o adaptive fusion | 35.34 | 30.04 | 39.01 | 30.74 | 43.29 | 53.85 |
| | (2.4). w/o joint training | 32.73 | 27.09 | 37.62 | 29.72 | 41.64 | 52.73 |
| | (2.5). w/o ExID | 34.99 | 29.49 | 38.42 | 30.63 | 42.42 | 53.24 |




Ablation study and parameter analysis for MoMoK

| Modality | Outperforming Predictions |
|---|---|
| Structure | RouteJunction, PartOf, Nearest City, Publisher, PrimeMinister, ComputingPlatform, LargestCity |
| Image | SisterStation, League, Parent, HubAirport, Company, Owner, Capital |
| Text | SisterStation, Publisher, Head, FederalState, Parent, ComputingPlatform, CountrySeat |
| Joint | GoverningBody, PartOf, Creator, Company, ComputingPlatform, RegionServed |
| Full Model | SisterStation, RouteJunction, GoverningBody, Publisher, PartOf, FederalState, ComputingPlatform, Parent |

Case study for the best predicted relations from each modality


Attention weights visualization for each experts under certain relation contexts.