Distance-Adaptive Quaternion Knowledge Graph Embedding with Bidirectional Rotation

Weihua Wang[1,2,3,*], Qiuyu Liang[1], Feilong Bao[1,2,3], Guanglai Gao[1,2,3]

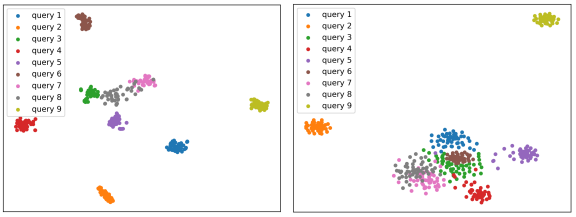
[1]College of Computer Science, Inner Mongolia University, Hohhot, China
[2]National and Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot, China
[3]Inner Mongolia Key Laboratory of Multilingual Artificial Intelligence Technology, Hohhot, China

Abstract

Quaternion contains one real part and three imaginary parts, which provided a more expressive hypercomplex space for learning knowledge graph. Existing quaternion embedding models measure the plausibility of a triplet through either semantic matching or geometric distance scoring functions. However, it appears that semantic matching diminishes the separability of entities, while the distance scoring function weakens the semantics of entities. To address this issue, we propose a novel quaternion knowledge graph embedding model. Our model combines semantic matching with the geometric distance of entities to better measure the plausibility of triplets. Specifically, in the quaternion space, we perform a right rotation on head entity and a reverse rotation on tail entity to learn rich semantic features. We then utilize distance-adaptive translations to learn geometric distance between entities. Furthermore, we provide mathematical proofs to demonstrate our model can handle complex logical relationships. Extensive experimental results and analyses show our model significantly outperforms previous models on well-known knowledge graph completion benchmark datasets. Our code is available at https://github.com/llqy123/DaBR.

1 Introduction

Knowledge graphs (KGs) (Liang et al., 2024a) are powerful tools for representing valid factual triplets by capturing entities and their relationships in a graphical format. Owing to the well-structured of graphs, KGs are often used for various Natural Language Processing tasks, such as question answering (Mendes et al., 2024; Faldu et al., 2024), entity alignment (Wang et al., 2024a,b), KG-based recommendation (Liang et al., 2024c) and KG enhanced Large Language Model (Wen et al., 2024).



(a) QuatE (b) TransERR

Figure 1: The visualization embedding of QuatE and TransERR models after 100 epochs training. Points in the same color represent tail entities that have the same (h_r, r_j) (query) context.

However, KGs are usually incomplete and the incompleteness limits their application. As an effective tool for predicting missing facts, knowledge graph completion (KGC) has received considerable attention from researchers. Typically, researchers transform KGC tasks into knowledge graph embeddings (KGEs). KGE refers to learning representations of entities and relations in a low-dimensional space while preserving the graph's inherent structure and semantic properties. In this representation space, a scoring function can be defined to measure the plausibility of each triplet, where valid triplets should receive higher scores than these invalid ones.

Quaternion contains one real part and three imaginary parts, which providing a more expressive space for learning embeddings of entities and relations. Rotation in the quaternion space is often used to model the KGs. For example, QuatE (Zhang et al., 2019) learns semantic information about entities by treating relations as rotations from head entities to tail entities. TransERR (Li et al., 2024) encodes the KG by rotating the head and tail entities with their corresponding unit quaternions. These models use either semantic matching or distance scoring functions to measure the plausibility of the triplet, respectively. However, it appears

*Corresponding Author. Email: wangwh@imu.edu.cn.

Proceedings of the 31st International Conference on Computational Linguistics, pages 4219–4231
January 19–24, 2025.

that semantic matching diminishes the separability of entities, while the distance scoring function weakens the semantics of entities. For example, we visualized the results for the same query in Figure 1 [1]. Specifically, as shown in Figure 1, we observe that QuatE model overlaps some queries when using semantic matching as a scoring function. The entities of TransERR using the distance scoring function are also indistinguishable from each query.

To address this issue, we propose a **D**istance-**a**daptive quaternion knowledge graph embedding with **B**idirectional **R**otation model, named as **DaBR**. Our model combines semantic matching with the geometric distance of entities to better measure the plausibility of triplets. Specifically, in the quaternion space, we perform a right rotation on the head entity and a reverse rotation on the tail entity to learn rich semantic features. This process is called bidirectional rotation. We conducted extensive experiments on multiple well-known benchmark datasets for knowledge graph completion task. The experimental results and analyses demonstrated the effectiveness and robustness of our model.

Our contributions are summarized as follows:

- We propose performing a right rotation on the head entity and a reverse rotation on the tail entity to learn rich semantic features.
- We propose learning the embedding distance between entities by incorporating distance adaptive translations.
- We provide mathematical proofs to demonstrate that our model can handle rich logical relationships.
- Extensive experiments show that our model provides consistent and significant improvements over previous models in most metrics.

2 Related Work

For KGE models, the design of the scoring function directly affects these models' performance and effectiveness. Based on the calculation methods of scoring functions in previous models, KGE scoring functions can mainly be categorized into semantic matching- and geometric distance-based.

Semantic matching. Semantic matching scoring functions capture the interactions between entities and relations through inner products on embedding vectors. The hypothesis is that entities connected by relations are close to each other in the semantic space. For example, QuatE (Zhang et al., 2019) obtains semantic information about entities through the Hamiltonian rotation of the head entity on the relation in quaternion space. DualE (Cao et al., 2021) further enhances QuatE to model knowledge graphs in dual quaternion space. QuatRE (Nguyen et al., 2022) associates each relation with two relation-aware rotations, which are used to rotate the quaternion embeddings of the head and tail entities, respectively. ConvQE (Liang et al., 2024d) investigates the potential of quaternion convolution in knowledge graph embedding.

A common feature of these models is the computation of the inner product between the head entity and the tail entity after a relation transformation. However, these models overlook the geometric distance properties between entities in the knowledge graph, which leads to distorted embeddings of the learned entities.

Geometric distance. Geometric distance scoring functions assess the plausibility of triplets by calculating the distances between embedding vectors in the representation space. The goal of this scoring function is to keep the head/tail entity vector closer to the tail/head entity vector after being transformed through the relation vector. For example, TransE (Bordes et al., 2013), considered the first model to employ a geometric distance scoring function, assumes that triplets (h, r, t) in knowledge graphs should satisfy the expression $h+r \approx t$. However, TransE struggles with more complex relation types, such as one-to-many (1-to-N), many-to-one (N-to-1) and many-to-many (N-to-N).

To address this limitation, several models using distance-based scoring functions have been proposed. For example, Rotate3D (Gao et al., 2020) maps entities to a 3D space, defining the relation as a rotation from the head entity to the tail entity. Trans4E (Nayyeri et al., 2021) performs rotations and translations in a quaternion space. RotateCT (Dong et al., 2022) transforms entity coordinates and represents each relation as a rotation in complex space. Rotate4D (Le et al., 2023) employs two distinct rotational transformations to align the head embedding with the tail embedding. DCNE (Dong et al., 2024) maps entities to the dual complex number space, using rotations in the 2D space

[1] For more information about queries, see Section 6.4.

through the multiplication of dual complex numbers to represent relations. TransERR (Li et al., 2024) encodes knowledge graphs by rotating the head and tail entities with their corresponding unit quaternions.

A common feature of these models is that the plausibility of the triplets is evaluated by calculating the distance between the head entity and the tail entity after transformation. However, these models do not consider information about entities within the semantic space, leading to performance degradation.

3 Preliminaries

This section begins with a definition of the knowledge graph completion task, followed by a brief background on quaternion algebra.

3.1 Knowledge Graph Completion

Knowledge graph completion is the task of predicting missing elements in a triplet (h, r, t). This task can be broken down into three sub-tasks: predicting the head entity $(?, r, t)$, predicting the relation $(h, ?, t)$, and predicting the tail entity $(h, r, ?)$. Following previous research, our work focuses on predicting the head $(?, r, t)$ and tail $(h, r, ?)$ entities. It is because relation information is needed in the training process.

3.2 Quaternion Algebra

The quaternion extends the complex number system to four dimensions. In n-dimensional quaternion space $\mathbb{Q}^n$, a quaternion $\mathbf{p} \in \mathbb{Q}^n$ consists of one real component and three imaginary components. It can be formalized as: $\mathbf{p} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, where $a, b, c, d \in \mathbb{R}^{\frac{n}{4}}$ are real numbers and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are imaginary units. The imaginary part satisfies the Hamilton's rules (Hamilton, 1844): $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$.

Addition. Given two quaternions $\mathbf{p} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ and $\mathbf{q} = e + f\mathbf{i} + g\mathbf{j} + h\mathbf{k} \in \mathbb{Q}^n$, quaternion addition is defined as:

$$\mathbf{p} + \mathbf{q} = (a+e) + (b+f)\mathbf{i} + (c+g)\mathbf{j} + (d+h)\mathbf{k} \quad (1)$$

Norm. The normalization of quaternions $\| \mathbf{p} \| \in \mathbb{Q}^n$ can be defined by the following:

$$\| \mathbf{p} \| = \sqrt{a^2 + b^2 + c^2 + d^2}. \quad (2)$$

Inverse. The inverse of quaternions $\| \mathbf{p} \| \in \mathbb{Q}^n$ can be defined by the following:

$$\mathbf{p}^{-1} = \frac{\bar{\mathbf{p}}}{\| \mathbf{p} \|^2}, \bar{\mathbf{p}} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}, \quad (3)$$

where $\bar{\mathbf{p}} \in \mathbb{Q}^n$ is the conjugate of $\mathbf{p} \in \mathbb{Q}^n$.

Hamilton product. Given two quaternions $\mathbf{p}$ and $\mathbf{q}$. The quaternion rotation of these two quaternions can be performed by the Hamilton product:

$$\begin{aligned} \mathbf{p} \otimes \mathbf{q} = &(a \circ e - b \circ f - c \circ g - d \circ h) + \\ &(b \circ e + a \circ f + c \circ h - d \circ g)\mathbf{i} + \\ &(c \circ e + a \circ g + d \circ f - b \circ h)\mathbf{j} + \\ &(d \circ e + a \circ h + b \circ g - c \circ f)\mathbf{k}, \end{aligned} \quad (4)$$

where $\circ$ denotes the element-wise product.

4 Methodology

In this section, we describe our model in detail, which consists of two main parts:

- **Bidirectional rotation**: Performing a right rotation on the head entity and a reverse rotation on the tail entity to learn the rich semantic features.
- **Distance-adaptation**: Incorporating a distance adaptive translation to learn the geometric distance between entity embeddings.

4.1 Symbol Description

A knowledge graph $\mathcal{G} =: \{(h, r, t)\} \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ is a collection of triplet, where $\mathcal{E}$ and $\mathcal{R}$ are the entity set and relation set. $|\mathcal{E}|$ and $|\mathcal{R}|$ represent the number of entities and relations, respectively. Given a triplet (h, r, t), the embeddings of head entity $\mathbf{h}$, relation $\mathbf{r}$ and tail entity $\mathbf{t}$ can be represented by quaternions:

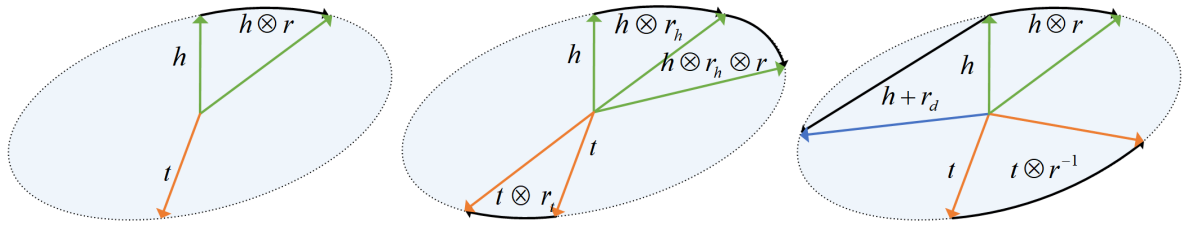
$$\begin{aligned} \mathbf{h} &= a_h + b_h\mathbf{i} + c_h\mathbf{j} + d_h\mathbf{k} \\ \mathbf{r} &= p + q\mathbf{i} + u\mathbf{j} + v\mathbf{k} \\ \mathbf{t} &= a_t + b_t\mathbf{i} + c_t\mathbf{j} + d_t\mathbf{k} \end{aligned} \quad (5)$$

4.2 Part One: Bidirectional Rotation

In Figure 2, we show the differences between our proposed bidirectional rotation and previous methods when modeling entity semantics. Specifically, QuatE (Figure 2(a)) performs a right rotation for head entity. QuatRE (Figure 2(b)) performs two times right rotation for head entity and a right rotation for tail entity. Our model (Figure 2(c)) performs a right rotation for head entity and a reverse rotation for tail entity.

We first normalize the relation quaternion $\mathbf{r}$ to a unit quaternion $\mathbf{r}^{\triangleleft}$ to eliminate the scaling effect by dividing by its norm (Equation 2):

$$\mathbf{r}^{\triangleleft} = \frac{\mathbf{r}}{\| \mathbf{r} \|} = \frac{p + q\mathbf{i} + u\mathbf{j} + v\mathbf{k}}{\sqrt{p^2 + q^2 + u^2 + v^2}}. \quad (6)$$



(a) QuatE (b) QuatRE (c) DaBR (ours)

Figure 2: The comparison of modeling entity semantics of QuatE, QuatRE and DaBR. These models learn the embeddings of knowledge graphs in quaternion spaces. $\otimes$ denotes the Hamilton product (Equation 4).

Then, the head entity $\mathbf{h}$ is right rotated using the relation $\mathbf{r}^{\triangleleft}$, i.e., the entity vector and the relation vector do a Hamilton product (Equation 4):

$$\mathbf{h}' = \mathbf{h} \otimes \mathbf{r}^{\triangleleft}. \tag{7}$$

Similarly, the inverse of the relation unit quaternion $\mathbf{r}^{\triangleleft}$ is used to make a reverse rotation of the tail entity $\mathbf{t}$:

$$\mathbf{t}' = \mathbf{t} \otimes \mathbf{r}^{\triangleleft -1}. \tag{8}$$

Since $\mathbf{r}^{\triangleleft}$ is a unit quaternion, we have:

$$\mathbf{t}' = \mathbf{t} \otimes \mathbf{r}^{\triangleleft -1} = \mathbf{t} \otimes \bar{\mathbf{r}}^{\triangleleft}, \tag{9}$$

where $\bar{\mathbf{r}}^{\triangleleft}$ is the conjugate of $\mathbf{r}^{\triangleleft}$.

Therefore, the scoring function $s(h, r, t)$ for the bidirectional rotation modeling entity semantics is defined by:

$$s(h, r, t) = \mathbf{h}' \circ \mathbf{t}' = \mathbf{h} \otimes \mathbf{r}^{\triangleleft} \circ \mathbf{t} \otimes \bar{\mathbf{r}}^{\triangleleft}, \tag{10}$$

4.3 Part Two: Distance-Adaptation

As shown in Figure 2, the previous QuatE (Figure 2(a)) and QuatRE (Figure 2(b)) can only learn the semantic information of an entity but ignore the geometric distance attribute of an entity. Our DaBR effectively addresses this limitation by adding a distance-adaptation (Figure 2(c)).

Therefore, to model the geometric distance information, we initialize a distance-adaptive relation embedding $\mathbf{r_d} = p_d + q_d\mathbf{i} + u_d\mathbf{j} + v_d\mathbf{k}$. Finally, the geometric distance part scoring function $d(h, r, t)$ is defined as:

$$d(h, r, t) = \| \mathbf{h} + \mathbf{r_d} - \mathbf{t} \|_1, \tag{11}$$

where $\| \cdot \|_1$ represents the ℓ_1 norm. Despite its simplicity, we find that the proposed method is effective enough in providing distance information for our model.

4.4 Scoring Function

After obtaining the scoring functions for modeling entity semantics and entity geometric distances, respectively. We fuse these scoring functions into a new scoring function for model training:

$$\begin{aligned} \phi(h, r, t) &= s(h, r, t) + \lambda d(h, r, t) \\ &= \mathbf{h} \otimes \mathbf{r}^{\triangleleft} \cdot \mathbf{t} \otimes \bar{\mathbf{r}}^{\triangleleft} + \lambda \| \mathbf{h} + \mathbf{r_d} - \mathbf{t} \|_1, \end{aligned} \tag{12}$$

where $s(h, r, t)$ represents the semantic matching scoring function, $d(h, r, t)$ represents the geometric distance scoring function, and $\lambda \in \mathbb{R}$ is an adaptive parameter that learned by our model.

4.5 Loss Function

Following Trouillon et al. (2016), we formulate the task as a classification problem, and the model parameters are learned by minimizing the following regularized logistic loss:

$$\begin{aligned} \mathcal{L} = \sum_{r(h,t) \in \Omega \cup \Omega^-} & \log(1 + \exp(-Y_{hrt}\phi(h, r, t))) \\ & + \eta_1 \| \mathbf{E} \|_2^2 + \eta_2 \| \mathbf{R} \|_2^2, \end{aligned} \tag{13}$$

where $\mathbf{E}$ and $\mathbf{R}$ denote the embedding of all entities and relations. Here we use the ℓ_2 norm with regularization rates η_1 and η_2 to regularize E and R, respectively. Ω^- is sampled from the unobserved set Ω' using uniform sampling. $Y_{hrt} \in \{-1, 1\}$ represents the corresponding label of the triplet (h, r, t).

4.6 Discussion

As described in Chami et al. (2020), there are complex logical relationships (such as symmetry, antisymmetry, inversion and composition relationships) in the knowledge graph. In this part, we analyze the ability of our DaBR to infer these relationships.

Lemma 1 *DaBR can infer the symmetry relationship pattern. (See proof in Appendix A.1)*

SF	Model	WN18RR					FB15k-237				
		MR(↓)	MRR	H@10	H@3	H@1	MR(↓)	MRR	H@10	H@3	H@1
SM	TuckER (2019)	-	.470	.526	.482	.443	-	.358	.544	.394	.266
	QuatE (2019)	2314	.488	.582	.508	.438	87	.348	.550	.382	.248
	DualE (2021)	2270	.492	.584	.513	.444	91	.365	.559	.400	.268
	QuatRE (2022)	1986	.493	.592	.519	.439	88	.367	.563	.404	.269
	ConvQE (2024d)	-	.487	.563	.502	.447	-	.366	.551	.402	.273
GD	ATTH (2020)	-	.486	.573	.499	.443	-	.348	.540	.384	.252
	Rotate3D (2020)	3328	.489	.579	.505	.442	165	.347	.250	.543	.385
	Trans4E (2021)	1755	.469	.577	.487	.416	158	.332	.527	.366	.236
	RotateCT (2022)	3285	.492	.579	.507	.448	171	.347	.537	.382	.251
	Rotate4D (2023)	3167	.499	.587	.518	**.455**	181	.353	.547	.391	.257
	CompoundE (2023)	-	.491	.576	.508	.450	-	.357	.545	.393	.264
	HAQE (2024e)	-	.496	.584	.512	.450	-	.343	.535	.379	.247
	DCNE (2024)	3244	.492	.581	.510	.448	169	.354	.547	.393	.257
	FHRE (2024b)	-	.494	.563	.510	.450	-	.345	.528	.375	.255
	TransERR (2024)	1167	.501	.605	.520	.450	125	.360	.555	.396	.264
SG	**DaBR (ours)**	**899**	**.510**	**.622**	**.538**	.450	**83**	**.373**	**.572**	**.410**	**.274**

Table 1: Knowledge graph completion results on WN18RR and FB15k-237 datasets. Best results are in bold and second best results are underlined. **SF** indicates the scoring function, **SM** indicates semantic matching scoring function, **GD** indicates geometric distance scoring function, and **SG** indicates our semantic matching and geometric distance scoring function. "-" indicates that there is no result reported. The same settings apply to Table 2.

Lemma 2 *DaBR can infer the antisymmetry relationship pattern. (See proof in Appendix A.2)*

Lemma 3 *DaBR can infer the inversion relationship pattern. (See proof in Appendix A.3)*

Lemma 4 *DaBR can infer the composition relationship pattern. (See proof in Appendix A.4)*

5 Experiments

In this section, we first introduce the datasets, evaluation protocol, implementation details and baselines. Subsequently, we evaluate our model on four benchmark datasets.

Datasets. To verify the effectiveness and robustness of our model, we conducted extensive experiments on four standard knowledge graph completion datasets including WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova and Chen, 2015), WN18 (Bordes et al., 2013) and FB15k (Bordes et al., 2013). The WN18 and FB15k datasets are known to suffer from a data leakage problem, which causes models to easily inferred and consequently performing well on metrics. WN18RR and FB15k-237 were derived as subsets of WN18 and FB15k respectively. These datasets are designed to address data leakage concerns and thereby present a more realistic prediction task. The detailed statistics of the four standard datasets are shown in Appendix B.

Evaluation protocol. Similar to previous work (Zhang et al., 2019; Li et al., 2024), we employed the filtered evaluation setup described in reference (Bordes et al., 2013) to filter out real triplets during the evaluation process. This was done to avoid flawed evaluations. We used evaluation metrics encompassed Mean Rank (MR), Mean Reciprocity Rating (MRR) and Hits@n (n=1, 3 or 10). Where a smaller value on the MR indicates a better model. The final scoring model on the test set is derived from the model with the highest Hits@10 score on the validation set.

Implementation details. We conduct all our experiments on a single NVIDIA GeForce RTX 4090 with 24GB of memory. The ranges of the hyper-parameters for the grid search are set as follows: the embedding dimension (dim) is selected from {300, 400, 500}; the learning rate (lr) is chosen from {0.01, 0.02, 0.05, 0.1}; and the number of negative triplets sampled (neg) per training triplet is selected from {5, 10}. The regularization rates η_1 and η_2 are adjusted within {0.01, 0.05, 0.1, 0.5}. We create 100 batches of training samples for different datasets. We optimize the loss function by utilizing Adagrad (Duchi et al., 2011). All our hyper-parameters are provided in Appendix C.

It is worth noting that our models **do not** employ the training strategies of self-adversarial negative sampling (Sun et al., 2019) or N3 regularization with reciprocal learning (Lacroix et al., 2018).

Baselines. To verify the effectiveness of our model, we compared DaBR with several powerful baseline

SF	Model	WN18					FB15k				
		MR(↓)	MRR	H@10	H@3	H@1	MR(↓)	MRR	H@10	H@3	H@1
SM	TuckER (2019)	-	.953	.958	.955	**.949**	-	.795	.892	.833	.741
	QuatE (2019)	162	.950	.959	.954	.945	**17**	.782	**.900**	.835	.711
	DualE (2021)	156	.952	.962	.956	.946	21	.813	.896	.850	.766
	QuatRE (2022)	116	.939	.963	.953	.946	21	.808	.896	.851	.751
GD	Rotate3D (2020)	214	.951	.961	.953	.945	39	.789	.887	.832	.728
	Trans4E (2021)	175	.950	.960	.953	.944	47	.767	.892	.834	.681
	RotateCT (2022)	201	.951	.963	.956	.944	34	.794	.888	.834	.737
	Rotate4D (2023)	173	.952	.963	.956	.946	37	.790	.887	.831	.732
	DCNE (2024)	192	.952	.963	.955	.945	34	.798	.888	.835	.745
	TransERR (2024)	82	.953	.965	.957	.945	41	.815	.896	.848	.767
SG	**DaBR (ours)**	**56**	**.954**	**.966**	**.959**	.946	18	**.819**	**.900**	**.854**	**.769**

Table 2: Knowledge graph completion results on WN18 and FB15k datasets.

Model	WN18RR				FB15k-237				WN18				FB15k			
	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1	MRR	H@10	H@3	H@1
DaBR	**.510**	**.622**	**.538**	**.450**	**.373**	**.572**	**.410**	**.274**	**.954**	**.966**	**.959**	**.946**	**.819**	**.900**	**.854**	**.769**
Variant I	.505	.617	.532	.445	.370	.569	.404	.272	.953	.964	.956	.943	.816	.894	.844	.766
Variant II	.495	.580	.512	.445	.368	.566	.402	.270	.947	.960	.954	.937	.801	.890	.847	.751

Table 3: Ablation results for all datasets.

models, including both well-known and recently proposed ones with outstanding results. We divide these models according to the scoring function:

1) Semantic Matching: TuckER (Balazevic et al., 2019), QuatE (Zhang et al., 2019), DualE (Cao et al., 2021), QuatRE (Nguyen et al., 2022) and ConvQE (Liang et al., 2024d).

2) Geometric Distance: ATTH (Chami et al., 2020), Rotate3D (Gao et al., 2020), Trans4E (Nayyeri et al., 2021), RotateCT (Dong et al., 2022), Rotate4D (Le et al., 2023), CompoundE (Ge et al., 2023), HAQE (Liang et al., 2024e), DCNE (Dong et al., 2024), FHRE (Liang et al., 2024b) and TransERR (Li et al., 2024).

For a fair comparison, we report the optimal results for these baselines from the original papers.

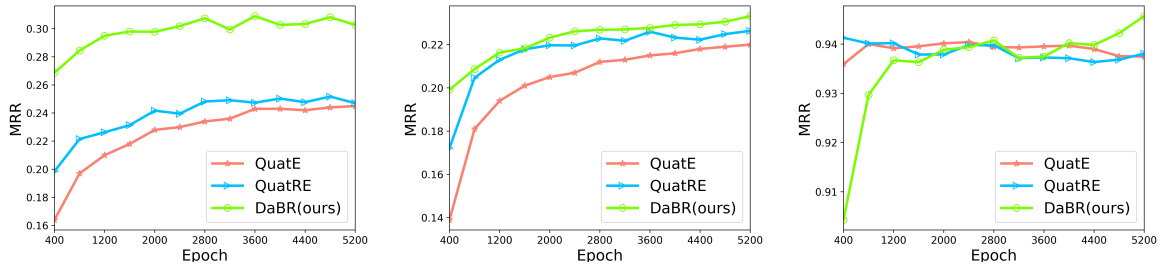
5.1 Main Results

The main results of our DaBR and the baselines for the WN18RR and FB15k-237 datasets are listed in Table 1. We categorize the baseline models into two main groups based on scoring functions, namely semantic matching and geometric distance scoring functions. The models based on **S**emantic **M**atching are listed in the upper part of the table, while the **G**eometric **D**istance based methods are listed in the lower part of the table. It is worth noting that our model's scoring function is the unique scoring function that simultaneously measures both **S**emantic and **G**eometric distances.

From Table 1 we can clearly see that our model achieves the best results on both datasets, except for the H@1 metric on the WN18RR dataset. Specifically, compared to the best performing of the semantic matching model, QuatRE, our model drops from 1986 to 899 on the MR metric and absolutely improves 3.4%, 5.0%, 3.6% and 2.5% on the MRR, H@10, H@3 and H@1 metrics on the WN18RR dataset. On the FB15k-237 dataset, our model decreases from 88 to 83 on the MR metrics, and absolutely improves on the MRR, H@10, H@3 and H@1 metrics by 1.6%, 1.5%, 1.4% and 1.8%.

Compared to the latest and best performance of the geometric distance model, TransERR, our model decreases from 1167 to 899 on the MR metric and achieves an absolute improvement of 1.8%, 2.8%, and 3.4% on the MRR, H@10 and H@3 metrics on the WN18RR dataset. On the FB15k-237 dataset, our model decreases from 125 to 83 on the MR metrics, and absolutely improves on the MRR, H@10, H@3 and H@1 metrics by 3.6%, 3.0%, 3.5% and 3.7%, respectively.

The KGC results on WN18 and FB15k datasets are shown in Table 2. The Table 2 illustrates our model superiority over any previous model on the FB15k dataset. On the WN18 dataset, our model achieves the best results on all metrics, except for the H@1 metric which achieves second place. In conclusion, our model not only achieves optimal results compared to semantic matching models, but also achieves competitive results compared to geometric distance models.



(a) 1-to-N (b) N-to-1 (c) N-to-N

Figure 3: MRR scores for QuatE, QuatRE and our DaBR models over 0 to 5200 training epochs.

6 Analysis

To demonstrate the superiority of our model, we have conducted in-depth analysis experiments from various aspects. The obtained experimental results and analysis are as follows:

6.1 Ablation Analysis

In this section, we aim to evaluate the efficacy of bidirectional rotation and distance-adaptation within our DaBR. We have designed the following model variants:

Variant I: We remove the rotation of the tail entity and keep the rotation of the head entity.

Variant II: We removed the distance-adaptation. The DaBR degenerates into a semantic matching model.

We show the results of the ablation experiments in Table 3. From the table, we can obtain the following conclusions: 1) The rotation of the tail entity and distance-adaptation are important parts of our model. 2) When our model removed the tail rotation, the model (i.e., *Variant I*) still achieved the best results compared to the models in Table 1 and Table 2. We attribute this to the fact that our model can measure both the semantics of entities and the embedding distance of entities. 3) When our model removed distance-adaptation, the model (i.e., *Variant II*) performance decreased dramatically on all datasets. It is worth noting that our model still achieves optimal results on most datasets compared to the semantic matching model on most datasets.

6.2 Parameter Comparison Analysis

To analyze the number of parameters compared to other models, we compared our DaBR with the best semantic matching model (QuatRE) and the best geometric distance model (TransERR). Given the same embedding dimension n, QuatRE and TransERR have $(|\mathcal{E}| \times n + 3 \times |\mathcal{R}| \times n)$ parameters, while our DaBR has $(|\mathcal{E}| \times n + 2 \times |\mathcal{R}| \times n)$ parameters, where $\mathcal{E}$ and $\mathcal{R}$ are the entity set and relation set. Compared to QuatRE and TransERR, our model achieves better results with fewer parameters.

6.3 Relationship Type Analysis

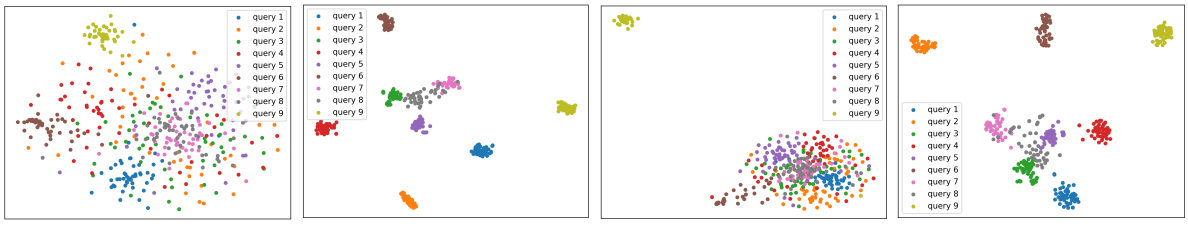
To explore the robustness of our model in the face of different relation types (one-to-many (1-to-N), many-to-one (N-to-1) and many-to-many (N-to-N)), we compared DaBR with QuatE and QuatRE in WN18R dataset. For the results of the QuatE and QuatRE, we reproduce these models following the hyper-parameter settings of their paper.

In accordance with the calculation rules set out in Bordes et al. (2013), the test set of WN18RR has been divided into three categories: 1-to-N, N-to-1 and N-to-N. The division results are shown in Appendix D, where η_h and η_t represent the average degree of head and tail entities, respectively.

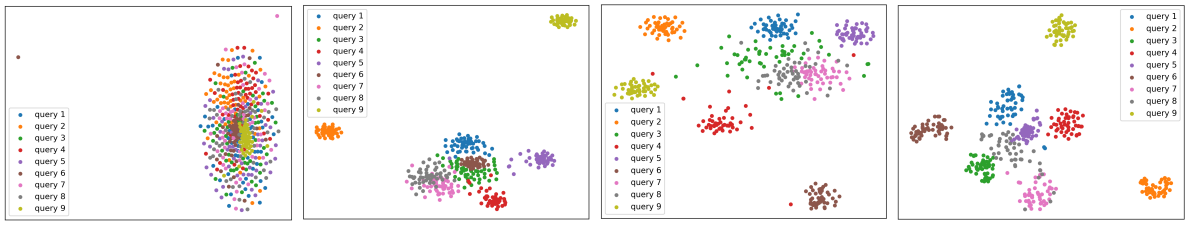
We show the MRR scores for the QuatE, QuatRE, and DaBR models for 0 to 5200 training epochs in Figure 3. This demonstrates the effectiveness of our model in modelling different types of relationships. In particular, the model is superior in dealing with 1-to-N relationship. "1-to-N" means that a head entity can form a fact triplet with multiple tail entities. We attribute this superior enhancement to the distance-adaptive embedding of our model.

6.4 Visualization Analysis

In this section, to explore the embedding results of our model after distance adaptive embedding, we visualize the the tail entity embeddings using t-SNE (van der Maaten and Hinton, 2008). Suppose (h_i, r_j) is a query where h_i and r_j are the head entity and the relation, respectively. If (h_i, r_j, t_k) is valid, the entity t_k is the answer to query (h_i, r_j).



(a) QuatE (epoch=1) (b) QuatE (epoch=100) (c) QuatRE (epoch=1) (d) QuatRE (epoch=100)



(e) TransERR (epoch=1) (f) TransERR (epoch=100) (g) DaBR (epoch=1) (h) DaBR (epoch=100)

Figure 4: Visualization of the embeddings of tail entities using t-SNE. A point represents a tail entity. Points in the same color represent tail entities that have the same (h_i, r_j) context.

We selected 9 queries in FB15k-237 dataset, each of which has 50 answers. For more details about the 9 queries, please refer to the Appendix E.

We then use t-SNE to visualize the semantic matching models QuatE and QuatRE, the geometric distance model TransERR, and our combined semantic and geometric distance DaBR to generate the answer embeddings for epoch 1 and epoch 100, respectively. Figure 4 shows the visualization results[2]. Each entity is represented by a 2D point and points in the same color represent tail entities with the same (h_i, r_j) context (i.e. query).

Specifically, our model (Figure 4(g)) in the first epoch have demonstrated better embedding compared to QuatE, QuatRE and TransERR. At epoch 100, our model (Figure 4(h)) show clear inter-cluster separability, with entities within each cluster (intra-cluster) being well-separated from one another.

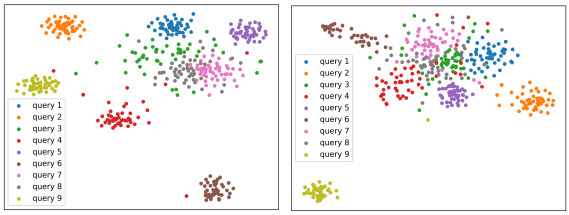
However, the semantic matching model QuatE (Figure 4(b)) and QuatRE (Figure 4(d)) heavily overlap entities within clusters despite inter-cluster separability. The geometric distance model TransERR (Figure 4(f)) clusters are indistinguishable from each other and entities within the clusters (intra-clusters) are distinguishable.

Table 4 summarizes our analysis above, which we attribute to the fact that our model combines semantic matching with entity geometric distance to better measure the plausibility of triplets.

[2]Refer to Appendix F for more visualization results.

Model	intra-cluster	inter-cluster
QuatE		✓
QuatRE		✓
TransERR	✓	
DaBR	✓	✓

Table 4: ✓ indicates a separable ability.



(a) DaBR (with) (b) DaBR (without)

Figure 5: DaBR with distance-adaptation and without.

6.5 Visualization Ablation Analysis

In Figure 5, we visualize that our model removes the distance adaptive embedding in the first epoch. We can find that the visualization without the distance adaptive embedding (Figure 5(b)) is worse than the with one (Figure 5(a)). By visualizing the ablation experiments, we can further illustrate the advantage of distance adaptive embedding.

7 Conclusion

We note that existing quaternion models based on semantic matching diminishes the separability of entities, while the distance scoring function weak-

ens the semantics of entities. To address this issue, we propose a novel quaternion knowledge graph embedding model. By combining semantic matching with entity geometric distance, our model provides a robust and comprehensive framework for knowledge graph embedding. We provide mathematical proofs to demonstrate our model can handle complex logical relationships. Visualization results show that our model can learn the geometric distance property between entities to achieve both inter-cluster and intra-cluster separability.

Limitations

The H@1 metric performance of our model on the WN18 and WN18RR datasets is not optimal. In addition, like most knowledge graph embedding models, our model is unable to predict new entities that do not exist in the training data.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No.62066033); Inner Mongolia Natural Science Foundation (Nos.2024MS06013, 2022JQ05); Inner Mongolia Autonomous Region Science and Technology Programme Project (Nos.2023YFSW0001, 2022YFDZ0059, 2021GG0158); We also thank all anonymous reviewers for their insightful comments.

References

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194, Hong Kong, China. Association for Computational Linguistics.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2787–2795. Curran Associates Inc.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. 2021. Dual quaternion knowledge graph embeddings. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):6894–6902.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6901–6914, Online. Association for Computational Linguistics.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press.

Yao Dong, Qingchao Kong, Lei Wang, and Yin Luo. 2024. Dual complex number knowledge graph embeddings. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5391–5400, Torino, Italia. ELRA and ICCL.

Yao Dong, Lei Wang, Ji Xiang, Xiaobo Guo, and Yuqiang Xie. 2022. RotateCT: Knowledge graph embedding by rotation and coordinate transformation in complex space. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4918–4932, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12(null):2121–2159.

Prayushi Faldu, Indrajit Bhattacharya, and Mausam . 2024. RetinaQA: A robust knowledge base question answering model for both answerable and unanswerable questions. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6643–6656, Bangkok, Thailand. Association for Computational Linguistics.

Chang Gao, Chengjie Sun, Lili Shan, Lei Lin, and Mingjiang Wang. 2020. Rotate3D: Representing relations as rotations in three-dimensional space for knowledge graph embedding. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, page 385–394, New York, NY, USA. Association for Computing Machinery.

Xiou Ge, Yun Cheng Wang, Bin Wang, and C.-C. Jay Kuo. 2023. Compounding geometric operations for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6947–6965, Toronto, Canada. Association for Computational Linguistics.

Wm R Hamilton. 1844. Theory of quaternions. *Proceedings of the Royal Irish Academy (1836-1869)*, 3:1–16.

Timothee Lacroix, Nicolas Usunier, and Guillaume Obozinski. 2018. Canonical tensor decomposition for knowledge base completion. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2863–2872. PMLR.

Thanh Le, Huy Tran, and Bac Le. 2023. Knowledge graph embedding with the special orthogonal group in quaternion space for link prediction. *Knowledge-Based Systems*, 266:1–26.

Jiang Li, Xiangdong Su, Fujun Zhang, and Guanglai Gao. 2024. TransERR: Translation-based knowledge graph embedding via efficient relation rotation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16727–16737, Torino, Italia. ELRA and ICCL.

Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. 2024a. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20.

Qiuyu Liang, Weihua Wang, Feilong Bao, and Guanglai Gao. 2024b. Fully hyperbolic rotation for knowledge graph embedding. In *2024 - 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain*, pages 1615–1622. IOS Press.

Qiuyu Liang, Weihua Wang, Lei Lv, and Feilong Bao. 2024c. Knowledge graph-enhanced recommendation with box embeddings. In *Chinese Computational Linguistics*, pages 274–288.

Qiuyu Liang, Weihua Wang, Jie Yu, and Feilong Bao. 2024d. Effective knowledge graph embedding with quaternion convolutional networks. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 183–196. Springer.

Qiuyu Liang, Weihua Wang, Jie Yu, and Feilong Bao. 2024e. Hierarchy-aware quaternion embedding for knowledge graph completion. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Renê Mendes, Dimas Oliveira, and Victor Garcia. 2024. Application of generative AI as an enterprise wikibase knowledge graph Q&A system. In *Proceedings of the 1st Workshop on Knowledge Graphs and Large Language Models (KaLLM 2024)*, pages 35–42, Bangkok, Thailand. Association for Computational Linguistics.

Mojtaba Nayyeri, Gokce Muge Cil, Sahar Vahdati, Francesco Osborne, Mahfuzur Rahman, Simone Angioni, Angelo Salatino, Diego Reforgiato Recupero, Nadezhda Vassilyeva, Enrico Motta, and Jens Lehmann. 2021. Trans4E: Link prediction on scholarly knowledge graphs. *Neurocomputing*, 461:530–542.

Dai Quoc Nguyen, Thanh Vu, Tu Dinh Nguyen, and Dinh Phung. 2022. QuatRE: Relation-aware quaternions for knowledge graph embeddings. In *Companion Proceedings of the Web Conference 2022*, WWW '22, page 189–192, New York, NY, USA. Association for Computing Machinery.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66. Association for Computational Linguistics.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 2071–2080. JMLR.org.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Cunda Wang, Weihua Wang, Qiuyu Liang, Feilong Bao, and Guanglai Gao. 2024a. Unifying dual-space embedding for entity alignment via contrastive learning. *arXiv preprint arXiv:2412.05028*.

Cunda Wang, Weihua Wang, Qiuyu Liang, Jie Yu, and Guanglai Gao. 2024b. Gsea: Global structure-aware graph neural networks for entity alignment. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 187–199. Springer.

Yilin Wen, Zifeng Wang, and Jimeng Sun. 2024. MindMap: Knowledge graph prompting sparks graph of thoughts in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10370–10388, Bangkok, Thailand. Association for Computational Linguistics.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA. Curran Associates Inc.

Appendix

A Proof

Given $\mathbf{h} = a_h + b_h\mathbf{i} + c_h\mathbf{j} + d_h\mathbf{k}, \mathbf{r} = p + q\mathbf{i} + u\mathbf{j} + v\mathbf{k}, \mathbf{t} = a_t + b_t\mathbf{i} + c_t\mathbf{j} + d_t\mathbf{k}$, where $\mathbf{r}$ is a unit quaternion after normalization operation. We can make $\lambda = 0$ and then our scoring function can be simplified as follows:

$$\begin{aligned}
&\phi(h,r,t) = \mathbf{h} \otimes \mathbf{r} \cdot \mathbf{t} \otimes \bar{\mathbf{r}} \\
&= [(a_h \circ p - b_h \circ q - c_h \circ u - d_h \circ v) \\
&+ (a_h \circ q + b_h \circ p + c_h \circ v - d_h \circ u)\mathbf{i} \\
&+ (a_h \circ u - b_h \circ v + c_h \circ p + d_h \circ q)\mathbf{j} \\
&+ (a_h \circ v + b_h \circ u - c_h \circ q + d_h \circ p)\mathbf{k}] \\
&\cdot [(a_t \circ p + b_t \circ q + c_t \circ u + d_t \circ v) \\
&+ (-a_t \circ q + b_t \circ p - c_t \circ v + d_t \circ u)\mathbf{i} \\
&+ (-a_t \circ u + b_t \circ v + c_t \circ p - d_t \circ q)\mathbf{j} \\
&+ (-a_t \circ v - b_t \circ u + c_t \circ q + d_t \circ p)\mathbf{k}]
\end{aligned} \tag{14}$$

where $\otimes$ is the Hamilton product, $\circ$ denotes the element-wise product, and "$\cdot$" is the inner product.

A.1 Proof of Symmetry pattern

In order to prove the symmetry pattern, we need to prove the following equality:

$$\mathbf{h} \otimes \mathbf{r} \cdot \mathbf{t} \otimes \bar{\mathbf{r}} = \mathbf{t} \otimes \mathbf{r} \cdot \mathbf{h} \otimes \bar{\mathbf{r}}. \tag{15}$$

The symmetry property of DaBR can be proved by setting the imaginary parts of $\mathbf{r}$ to zero.

A.2 Proof of Antisymmetry pattern

In order to prove the antisymmetry pattern, we need to prove the following inequality when imaginary components are nonzero:

$$\mathbf{h} \otimes \mathbf{r} \cdot \mathbf{t} \otimes \bar{\mathbf{r}} \neq \mathbf{t} \otimes \mathbf{r} \cdot \mathbf{h} \otimes \bar{\mathbf{r}}. \tag{16}$$

We expand the right term:

$$\begin{aligned}
&\mathbf{t} \otimes \mathbf{r} \cdot \mathbf{h} \otimes \bar{\mathbf{r}} \\
&= [(a_t \circ p - b_t \circ q - c_t \circ u - d_t \circ v) \\
&+ (a_t \circ q + b_t \circ p + c_t \circ v - d_t \circ u)\mathbf{i} \\
&+ (a_t \circ u - b_t \circ v + c_t \circ p + d_t \circ q)\mathbf{j} \\
&+ (a_t \circ v + b_t \circ u - c_t \circ q + d_t \circ p)\mathbf{k}] \\
&\cdot [(a_h \circ p + b_h \circ q + c_h \circ u + d_h \circ v) \\
&+ (-a_h \circ q + b_h \circ p - c_h \circ v + d_h \circ u)\mathbf{i} \\
&+ (-a_h \circ u + b_h \circ v + c_h \circ p - d_h \circ q)\mathbf{j} \\
&+ (-a_h \circ v - b_h \circ u + c_h \circ q + d_h \circ p)\mathbf{k}].
\end{aligned} \tag{17}$$

We can easily see that those two terms are not equal as the signs for some terms are not the same.

A.3 Proof of Inversion pattern

To prove the inversion pattern, we need to prove that:

$$\mathbf{h} \otimes \mathbf{r} \cdot \mathbf{t} \otimes \bar{\mathbf{r}} = \mathbf{t} \otimes \bar{\mathbf{r}} \cdot \mathbf{h} \otimes \bar{\mathbf{r}}^{-1}. \tag{18}$$

We expand the right term:

$$\begin{aligned}
&\mathbf{t} \otimes \bar{\mathbf{r}} \cdot \mathbf{h} \otimes \bar{\mathbf{r}}^{-1} \\
&= \mathbf{t} \otimes \bar{\mathbf{r}} \cdot \mathbf{h} \otimes \mathbf{r} \\
&= [(a_t \circ p + b_t \circ q + c_t \circ u + d_t \circ v) \\
&+ (-a_t \circ q + b_t \circ p - c_t \circ v + d_t \circ u)\mathbf{i} \\
&+ (-a_t \circ u + b_t \circ v + c_t \circ p - d_t \circ q)\mathbf{j} \\
&+ (-a_t \circ v - b_t \circ u + c_t \circ q + d_t \circ p)\mathbf{k}] \\
&\cdot [(a_h \circ p - b_h \circ q - c_h \circ u - d_h \circ v) \\
&+ (a_h \circ q + b_h \circ p + c_h \circ v - d_h \circ u)\mathbf{i} \\
&+ (a_h \circ u - b_h \circ v + c_h \circ p + d_h \circ q)\mathbf{j} \\
&+ (a_h \circ v + b_h \circ u - c_h \circ q + d_h \circ p)\mathbf{k}].
\end{aligned} \tag{19}$$

We can easily check the equality of these two terms. Since $\mathbf{r}$ is a unit quaternion, we have $\mathbf{r}^{-1} = \bar{\mathbf{r}}$.

A.4 Proof of Composition pattern

For composition relationships, we can get that:

$$\begin{aligned}
&(\mathbf{h} \otimes \mathbf{r_2}) \otimes \mathbf{r_3} \cdot (\mathbf{t} \otimes \bar{\mathbf{r}}_\mathbf{2}) \otimes \bar{\mathbf{r}}_\mathbf{3} \\
&= \mathbf{h} \otimes (\mathbf{r_2} \otimes \mathbf{r_3}) \cdot \mathbf{t} \otimes (\bar{\mathbf{r}}_\mathbf{2} \otimes \bar{\mathbf{r}}_\mathbf{3}) \\
&= \mathbf{h} \otimes \mathbf{r_1} \cdot \mathbf{t} \otimes \bar{\mathbf{r}}_\mathbf{1}
\end{aligned} \tag{20}$$

B Dataset statistics

The detailed statistics of the four standard datasets are shown in Table 6.

C Optimal hyper-parameters

Table 7 shows the optimal hyperparameter settings for our model on the four benchmark datasets. The optimal parameters come from the highest scores of our model on the validation dataset.

D Classification rules

The classification rules and classification results for WN18RR dataset in the Table 8.

E The queries in t-SNE visualization

In Table 5, we list the nine queries used in the t-SNE visualization (Section 6.4 in the main text). Note that a query is represented as $(h, r, ?)$, where h denotes the head entity and r denotes the relation.

F More visualization results

Figure 6 shows more visualization results.

Index	Query
1	(political drama, /media_common/netflix_genre/titles, ?)
2	(Academy Award for Best Original Song, /award/award_category/winners./award/award_honor/ceremony, ?)
3	(Germany, /location/location/contains, ?)
4	(Master's Degree, /education/educational_degree/people_with_this_degree./education/education/major_field_of_study, ?)
5	(broccoli, /food/food/nutrients./food/nutrition_fact/nutrient, ?)
6	(shooting sport, /olympics/olympic_sport/athletes./olympics/olympic_athlete_affiliation/country,?)
7	(synthpop, /music/genre/artists, ?)
8	(Italian American, /people/ethnicity/people, ?)
9	(organ, /music/performance_role/track_performances./music/track_contribution/role, ?)

Table 5: The queries in t-SNE visualizations.

Dataset	#Ent	#Rel	#Train	#Valid	#Test
WN18RR	$40k$	11	$86k$	$3k$	$3k$
FB15k-237	$14k$	237	$272k$	$17k$	$20k$
WN18	$40k$	18	$141k$	$5k$	$5k$
FB15k	$14k$	1345	$483k$	$50k$	$59k$

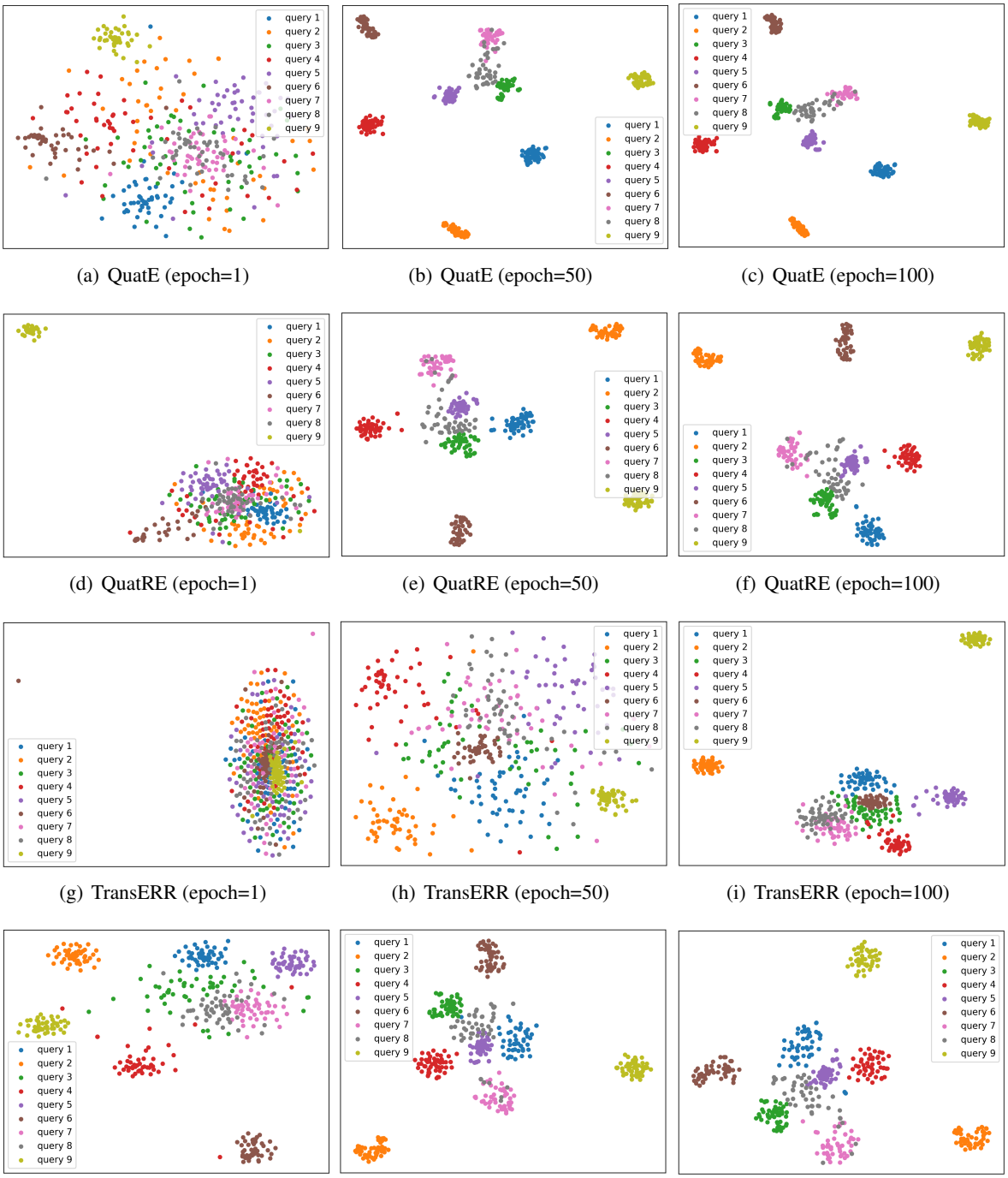
Table 6: Dataset statistics on four datasets.

Dataset	lr	neg	dim	η_1	η_2
WN18RR	0.1	5	500	0.5	0.01
FB15k-237	0.05	10	500	0.5	0.01
WN18	0.05	5	300	0.05	0.01
FB15k	0.02	10	400	0.05	0.01

Table 7: Optimal hyper-parameters for our DaBR on each dataset.

Category	η_h	η_t	#triplets
1-to-N	< 1.5	> 1.5	475
N-to-1	> 1.5	< 1.5	1487
N-to-N	> 1.5	> 1.5	1130

Table 8: Classification rules and classification results for WN18RR. The last column is the number after division.



(a) QuatE (epoch=1) (b) QuatE (epoch=50) (c) QuatE (epoch=100)

(d) QuatRE (epoch=1) (e) QuatRE (epoch=50) (f) QuatRE (epoch=100)

(g) TransERR (epoch=1) (h) TransERR (epoch=50) (i) TransERR (epoch=100)

(j) DaBR (epoch=1) (k) DaBR (epoch=50) (l) DaBR (epoch=100)

Figure 6: Visualization of the embeddings of tail entities using t-SNE. A point represents a tail entity. Points in the same color represent tail entities that have the same (h_r, r_j) context.