

RatE: Relation-Adaptive Translating Embedding for Knowledge Graph Completion

Hao Huang, Guodong Long*, Tao Shen, Jing Jiang and Chengqi Zhang

Australian AI Institute, School of Computer Science, FEIT, University of Technology Sydney

{hao.huang-4, tao.shen}@student.uts.edu.au

{guodong.long, jing.jiang, chengqi.zhang}@uts.edu.au

Abstract

Many graph embedding approaches have been proposed for knowledge graph completion via link prediction. Among those, translating embedding approaches enjoy the advantages of light-weight structure, high efficiency and great interpretability. Especially when extended to complex vector space, they show the capability in handling various relation patterns including symmetry, antisymmetry, inversion and composition. However, previous translating embedding approaches defined in complex vector space suffer from two main issues: 1) representing and modeling capacities of the model are limited by the translation function with rigorous multiplication of two complex numbers; and 2) embedding ambiguity caused by one-to-many relations is not explicitly alleviated. In this paper, we propose a relation-adaptive translation function built upon a novel weighted product in complex space, where the weights are learnable, relation-specific and independent to embedding size. The translation function only requires eight more scalar parameters each relation, but improves expressive power and alleviates embedding ambiguity problem. Based on the function, we then present our Relation-adaptive translating Embedding (RatE) approach to score each graph triple. Moreover, a novel negative sampling method is proposed to utilize both prior knowledge and self-adversarial learning for effective optimization. Experiments verify RatE achieves state-of-the-art performance on four link prediction benchmarks.

1 Introduction

A knowledge graph refers to a collection of interlinked entities, which is usually formatted as a set of triples. A triple is represented as a head entity linked to a tail entity by a relation, which is written as $(\text{head}, \text{relation}, \text{tail})$ or (h, r, t) . Large-scale knowledge graphs, such as Freebase (Bollacker et al., 2008) and WordNet (Miller, 1995), containing structured information, have been leveraged to support a broad spectrum of natural language processing (NLP) tasks, e.g., question answering (Hao et al., 2017), recommender system (Zhang et al., 2016), relation extraction (Min et al., 2013), etc. Nonetheless, the human-curated, real-world knowledge graphs often suffer from incompleteness or sparseness problem (Toutanova et al., 2015), which inevitably hurts the performance of downstream tasks. Hence, how to auto-complete knowledge graphs becomes a popular problem in both research and industry communities.

For this purpose, many light-weight graph embedding approaches (Bordes et al., 2013; Yang et al., 2015; Sun et al., 2019) have been proposed. Unlike costly graph neural networks (GNNs) (Schlichtkrull et al., 2018), these approaches use low-dimensional embeddings to represent the entities and relations, and capture their relationships via semantic matching or geometric distance. Specifically, the approaches with semantic matching, e.g., DistMult (Yang et al., 2015) and QuatE (Zhang et al., 2019), use a matching function $f(\mathbf{h}, \mathbf{r}, \mathbf{t})$ that operates on whole triple to directly derive its plausibility score. In contrast, the approaches with geometric distance, e.g., TransE (Bordes et al., 2013) and RotatE (Sun et al., 2019), first apply a translation function to head entity and relation for a new embedding in latent space and then measure a distance from the new embedding to tail entity, i.e., $f(\mathbf{h}, \mathbf{r}, \mathbf{t}) = -\|\mathbf{g}(\mathbf{h}, \mathbf{r}) - \mathbf{t}\|_p$. Empirically, the latter, namely *trans-based approach*, usually has higher efficiency and superior performance

*Corresponding author.

Type	Model	Score Function	Space	Sym.	Antisym.	Inv.	Comp.	Disambiguation
Semantic matching	DistMult	$\langle r, h, t \rangle$	\mathbb{R}^d	✓	✗	✗	✗	-
	ComplEx	$\text{Re}(\langle r, h, \bar{t} \rangle)$	\mathbb{C}^d	✓	✓	✓	✗	-
	QuatE	$h \otimes r^\triangleleft \cdot t$	\mathbb{H}^d	✓	✓	✓	✗	-
Trans-based	TransE	$- h + r - t $	\mathbb{R}^d	✗	✓	✓	✓	✗
	RotatE	$- h \circ r - t _1$	\mathbb{C}^d	✓	✓	✓	✓	✗
	RatE	$- h \odot_{W(r)} r - t _1$	\mathbb{C}^d	✓	✓	✓	✓	✓

Table 1: A brief comparison of semantic matching and trans-based graph embedding approaches, where a check mark denotes the model is equipped with the corresponding property. ‘‘Sym.’’, ‘‘Antisym.’’, ‘‘Inv.’’ and ‘‘Comp.’’ are abbreviations of relation patterns of symmetry, antisymmetry, inversion and composition respectively. For a trans-based graph embedding model, ‘‘Disambiguation’’ denotes whether the model explicitly handles embedding ambiguity problem as detailed in §2.5. And, $\langle \cdot \rangle$ denotes generalized dot product, \circ denotes element-wise (Hadamard) complex product, \otimes denotes element-wise Hamilton product, $^\triangleleft$ denotes normalization of a vector, and \odot_W denotes our proposed weighted product defined in Eq.(2).

on link prediction than the former. Based on translating process, it also offers better interpretability of the graph embeddings and relation modeling (Sun et al., 2019).

Recently, some trans-based graph embedding approaches, e.g., RotatE (Sun et al., 2019), go beyond real vector space. They represent the entities and relations in complex vector space, and define the translation function on complex vectors. Empowered by the properties of arithmetic operations (e.g., product) in complex space, the translation function can easily capture relation patterns of symmetry (e.g., *marriage*), antisymmetry (e.g., *father*), inversion(e.g., *hypernym* vs. *hyponym*) and composition (e.g., *mother* \wedge *husband* \rightarrow *father*). Compared to those defined in real vector space, these approaches improve model’s capability in handling a variety of relation patterns and achieve state-of-the-art performance.

Nevertheless, current trans-based graph embedding approaches with complex embeddings are vulnerable to the following two issues. On the one hand, although approaches solely in complex vector space are equipped with high interpretability for various relation patterns, they are limited by the expressive power of standard product/add of two complex numbers. To improve, QuatE (Zhang et al., 2019) introduces quaternion hypercomplex vector space with semantic matching, at the cost of both interpretability and computational overheads, but the improvement is still marginal. On the other hand, embedding ambiguity problem, which means different entities are assigned with similar embeddings, cannot be explicitly handled by existing trans-based approaches (e.g., TransE and RotatE). It is mainly caused by the propagation of applying a translation function to one-to-many relations for optimizing $\forall t = g(h, r)$.

To alleviate both issues above, we propose a novel **R**elation-**a**daptive **t**ranslating **E**mbedding (RatE) approach for knowledge graph completion. As an extension of the trans-based embedding approach RotatE, our proposed RatE inherits the capability to handle various relation patterns, and further presents a light-weight yet effective relation-adaptive translation function. Specifically, the function is composed of a novel element-wise *weighted product* defined in complex vector space, where the weights are learnable, relation-specific and independent to embedding dimension. Rather than rigorous complex number product in RotatE and QuatE, RatE provides a more flexible way – either the resulting real or imaginary part is a weighted sum of the product on every pair of numbers respectively from the two complex number arguments (i.e., real or imaginary part). Hence, RatE only requires *eight* more scalar parameters each relation than baseline RotatE, which is much less than the embedding dimension by one or two orders of magnitude. Through relation-adaptive translation function, the proposed approach empirically promotes the capacity of modeling translation process and embedding ambiguity problem, while preserves most interpretability to handle various relation patterns.

We also propose a novel local-cognitive negative sampling method, by integrating type-constraint training technique (Krompaß et al., 2015) with self-adversarial learning (Sun et al., 2019). The former leverages prior knowledge in graph during training and samples negative head (tail) entities from relation-specific domain (range), which is limited by the hard sampling criterion and suffers from graph sparseness. By comparison, the latter scores a certain number of uniformly-sampled negative samples based on current model, and uses the normalized scores as weights for the loss function. It hence depends heavily on an incompletely-trained model. Thus, we integrate them for their mutual benefits: besides using a self-adversarial loss, our method leverages prior knowledge to weaken the effect of current model.

Our main contributions are summarized in the following.

- We propose a trans-based graph embedding approach with a novel relation-adaptive translation function in complex vector space, which achieves a better trade-off between interpretability and representing capacity than previous approaches.
- We verify the model’s capability in alleviating embedding ambiguity problem caused by one-to-many relation pattern, from both theoretical and empirical perspectives.
- With a novel negative sampling method, we evaluate the proposed approach on four link prediction benchmark datasets, i.e., WN18, FB15k, WN18RR and FB15k-237, which shows state-of-the-art results among semantic matching and trans-based graph embedding approaches. The experimental codes are available at <https://github.com/Hhnro/RatE>.

2 Proposed Approach

This section begins with a definition of link prediction task for knowledge graph completion, followed by an introduction to a baseline RotatE. Then, we propose a novel relation-adaptive translation function to compose the final relation-adaptive translating embedding approach. Then, we present an efficient negative sampling method by integrating the merits of two previous sampling strategies. Lastly, we demonstrate the capability of our proposed model in alleviating embedding ambiguity problem.

2.1 Link Prediction

Formally, a knowledge graph $\mathcal{G} = \{\mathcal{E}, \mathcal{R}\}$ consists of a set of triples (h, r, t) , where $h, t \in \mathcal{E}$ are head and tail entities respectively while $r \in \mathcal{R}$ is the relation between them. Given a head h (or tail t) entity and a relation r , the goal of link prediction is to find the most accurate tail t (or head h) from \mathcal{E} to make the new triple (h, r, t) plausible in the knowledge graph \mathcal{G} . In a graph embedding approach, each entity/relation is assigned with an embedding vector, and a triple is denoted as $(\mathbf{h}, \mathbf{r}, \mathbf{t})$. To tackle link prediction, a scoring function $f(\mathbf{h}, \mathbf{r}, \mathbf{t})$ is presented to derive the plausibility score for each triple candidate. Especially in a trans-based approach, the score function is formulated as $f(\mathbf{h}, \mathbf{r}, \mathbf{t}) = -\|\mathbf{g}(\mathbf{h}, \mathbf{r}) - \mathbf{t}\|_p$ where $\mathbf{g}(\cdot)$ denotes a translation function.

2.2 Baseline: RotatE

RotatE is a state-of-the-art trans-based graph embedding approach in complex vector space. Motivated by Euler’s identity, its translating process is formulated as a relation-specific rotation of the head’s embedding vector. RotatE in complex space can be viewed as a natural extension of vanilla TransE in real vector space, aiming to support the relation pattern of symmetry. Specifically, RotatE represents both entities \mathcal{E} and relations \mathcal{R} in complex vector space \mathbb{C}^d , and defines relation’s embedding as a rotation by constraining the modulus of each dimension to 1. And its translation function $\mathbf{g}(\mathbf{h}, \mathbf{r})$ is simply fulfilled by a Hadamard product (i.e., element-wise, denoted as “ \circ ”) in complex vector space, i.e., $\mathbf{g}(\mathbf{h}, \mathbf{r}) = \mathbf{h} \circ \mathbf{r}$. Therefore, the scoring function in RotatE is written as

$$f(\mathbf{h}, \mathbf{r}, \mathbf{t}) = -\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_1, \text{ where } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^d \text{ and } \forall |r_i| = 1. \quad (1)$$

Note, the p -norm of a complex vector \mathbf{v} is defined as $\|\mathbf{v}\|_p = \sqrt[p]{\sum |v_i|^p}$.

2.3 Relation-Adaptive Translating Embedding

Based on the baseline, we propose a trans-based graph embedding approach, named as **Relation-adaptive translating Embedding** (RatE). It extends complex number product to a novel *weighted product* in complex space, where the weights are learnable and relation-specific. The weighted product is defined as

$$\begin{aligned} o = u \otimes_{\mathbf{W}} v &= (a + bi) \otimes_{\mathbf{W}} (c + di) = \mathbf{W}_{1,:} \mathbf{s}^{(u,v)} + \mathbf{W}_{2,:} \mathbf{s}^{(u,v)} i, \\ \text{where, } o, u, v \in \mathbb{C}, \mathbf{W} \in \mathbb{R}^{2 \times 4} \text{ and } \mathbf{s}^{(u,v)} &= [ac; ad; bc; bd] \in \mathbb{R}^4. \end{aligned} \quad (2)$$

Here, \mathbf{W} denotes a learnable weight matrix and will be updated during training for a specific target. Standard complex number product is its special case when $\mathbf{W} = [[1, 0, 0, -1]; [0, 1, 1, 0]]$. Hence, empowered by the learnable weights, the weighted product promotes the ability to implicitly capture arithmetic or geometrical relationships in complex space when adapted into a data-driven neural model.

Then, the proposed weighted product is readily integrated with RotatE to compose a novel relation-adaptive translation function. That is

$$\tilde{\mathbf{t}} := g(\mathbf{h}, \mathbf{r}) = \mathbf{h} \odot_{\mathbf{W}(r)} \mathbf{r}, \text{ where, } \forall i : \tilde{t}_i = h_i \otimes_{\mathbf{W}(r)} r_i, |r_i| = 1. \quad (3)$$

$\mathbf{h}, \mathbf{r} \in \mathbb{C}^d$ are the embeddings of head entity and relation respectively, and $\odot_{\mathbf{W}(r)}$ denotes *element-wise weighted product* where the weights are specified for each relation $r \in \mathcal{R}$. Based on this translation function, we formulate the score function of relation-adaptive translating embedding as

$$s^{(h,r,t)} := f(\mathbf{h}, \mathbf{r}, \mathbf{t}) = -||\mathbf{h} \odot_{\mathbf{W}(r)} \mathbf{r} - \mathbf{t}||_1, \quad (4)$$

where $s^{(h,r,t)} \in \mathbb{R}$ is the resulting score of the triple (h, r, t) to measure its plausibility. As both the graph embeddings and the translation function are defined in complex vector space and learnable during training, our proposed RatE is a generic formulation of previous trans-based approaches. In other words, the approaches like RotatE and TransE are special cases of RatE, so our approach makes the best of deep neural network and promotes the representing capacity of translating paradigm. This is achieved by increasing only *eight* learnable parameters for each relation, which are fewer than the relation's embedding size by one or two orders of magnitude. Moreover, besides handling the four relation patterns (i.e., symmetry, antisymmetry, inversion and composition), the proposed RatE also reduces the effect of embedding ambiguity (detailed at the end of this section). It is also noteworthy that although the integration above is based on RotatE, the proposed weighted product is compatible with any complex or hypercomplex embedding approach (e.g., QuatE).

2.4 Negative Sampling and Optimization

The way to conduct negative sampling can significantly affect the performance of a graph embedding approach (Cai and Wang, 2018; Sun et al., 2019) because contrasting a challenging negative sample against the corresponding positive one is more effective for learning structured knowledge. Formally, given an arbitrary correct triple $x = (h, r, t) \in \mathcal{G}^{(tr)}$, negative sampling aims at corrupting its either head or tail entity to get a wrong triple $x' = (h', r, t)$ or (h, r, t') , where $x' \notin \mathcal{G}^{(tr)}$. $\mathcal{G}^{(tr)}$ denotes the knowledge graph to train an embedding model. Note, we only exhibit tail corruption for a clear elaboration in the following, and head corruption is also considered in our implementation.

We first introduce two popular sampling strategies in the following. Type-constraint training technique (Krompaß et al., 2015) presents a new link prediction setting based on local closed-world assumptions – the entities to corrupt a triple only come from a relation-specific entity set during both training and test. We only take this idea in training phase to introduce prior knowledge and provide strong distractors. Particularly, for a triple (h, r, t) , the candidate set of tail corruptions is

$$\mathcal{E}^{(h,r,t)} = \{t' \in \mathcal{E} | \exists e \in \mathcal{E} : (e, r, t') \in \mathcal{G}^{(tr)} \wedge (h, r, t') \notin \mathcal{G}^{(tr)}\}. \quad (5)$$

However, sampling only in this set, $\mathcal{E}^{(h,r,t)}$, suffers from not only graph sparseness by local closed-world assumptions but also information loss of other corrupting entities. The other entities are denoted as

$$\bar{\mathcal{E}}^{(h,r,t)} = \{t' \in \mathcal{E} | t' \notin \mathcal{E}^{(h,r,t)} \wedge (h, r, t') \notin \mathcal{G}^{(tr)}\}. \quad (6)$$

In contrast, self-adversarial negative sampling (Sun et al., 2019) applies triple scoring function to a certain number of uniformly-sampled wrong triples, and each $f(\mathbf{h}, \mathbf{r}, \mathbf{t}')$ represents its difficulty to current embedding model. It then uses the normalized scores as the weights in loss function to perform a self-adversarial training. However, this sampling strategy depends heavily on current embedding model.

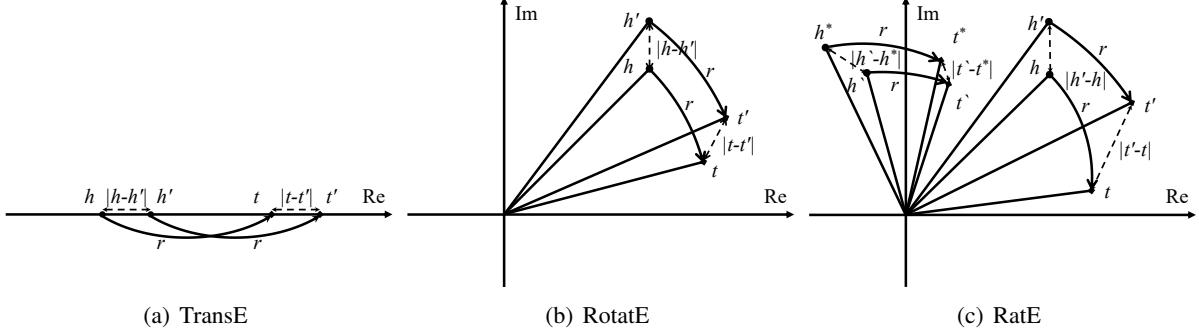


Figure 1: Toy examples – applying translation functions of TransE, RotatE and RatE to (h_i, r_i) for the resulting t_i . Note that 1) dimension index i is omitted, and 2) TransE is defined in real space whereas RotatE/RatE is defined in complex space.

Then, we propose a novel local-cognitive negative sampling method by integrating them to complement each other. Our integration is non-trivial, where a dynamic coefficient¹ $\gamma \in [0, 1]$ is used to control the proportion of negative samples from $\mathcal{E}^{(h,r,t)}$ or $\bar{\mathcal{E}}^{(h,r,t)}$. In particular, a certain number n of wrong triples is first sampled for each triple $x = (h, r, t) \in \mathcal{G}^{(tr)}$. To achieve this, we conduct a uniform sampling individually in $\mathcal{E}^{(h,r,t)}$ and $\bar{\mathcal{E}}^{(h,r,t)}$, which respectively produce \mathcal{N} containing γn samples and $\bar{\mathcal{N}}$ containing $(1 - \gamma)n$ samples. Then we optimize the proposed embedding model by minimizing

$$\mathcal{L} = \mu \|\mathbf{W}^{(r)}\|_1 - \log \sigma(\lambda + f(\mathbf{h}, \mathbf{r}, \mathbf{t})) - \sum_{(h,r,t') \in \mathcal{N} \cup \bar{\mathcal{N}}} \beta^{(h,r,t')} \log \sigma(-f(\mathbf{h}, \mathbf{r}, \mathbf{t}') - \lambda), \quad (7)$$

$$\text{where } \beta^{(h,r,t')} = \exp f(\mathbf{h}, \mathbf{r}, \mathbf{t}') / \sum_{(h,r,t'') \in \mathcal{N} \cup \bar{\mathcal{N}}} \exp f(\mathbf{h}, \mathbf{r}, \mathbf{t}''). \quad (8)$$

μ is weight decay of L1 regularization and set to 0.01 without tuning. Lastly, we update the coefficient γ at the end of every training epoch by

$$\gamma \leftarrow \frac{1}{|\mathcal{G}^{(tr)}|} \sum_{\mathcal{G}^{(tr)}} 1 / \left(1 + \frac{\sum_{(h,r,t') \in \bar{\mathcal{N}}} \exp f(\mathbf{h}, \mathbf{r}, \mathbf{t}') / |\bar{\mathcal{N}}|}{\sum_{(h,r,t') \in \mathcal{N}} \exp f(\mathbf{h}, \mathbf{r}, \mathbf{t}') / |\mathcal{N}|} \right). \quad (9)$$

Here γ inclines to the candidate set with more challenging negative samples, which is determined by all wrong triples sampled in the previous epoch. In summary, our sampling method employs a self-adversarial training loss, and leverages prior knowledge to weaken the effect of current model.

2.5 Embedding Disambiguation

Embedding ambiguity here refers to similar embeddings assigned to distinct entities. In a trans-based graph embedding approach, it is usually caused by one-to-many (i.e., a kind of non-injective) relations in knowledge graphs. Specifically, given a set of triples $\{(h, r, t^1), \dots, (h, r, t^M)\}$ as an example of one-to-many relations, invoking a translation function directly defined in real or complex space makes the model optimize toward $\forall t^j = g(\mathbf{h}, \mathbf{r})$ and inevitably results in similar tail embeddings. Because one-to-many relations are ubiquitous in a knowledge graph, e.g., *has_part* in WordNet, the embedding ambiguity problem will deteriorate and propagate through the graph. Fortunately, the proposed RatE is able to alleviate this problem by cutting off the propagation.

To intuitively demonstrate RatE’s capability in embedding disambiguation by stopping the propagation, we respectively illustrate toy examples of TransE, RotatE, and our proposed RatE in Figure 1. It is observed that given two head entities with similar embeddings, their similarity will be preserved in corresponding tail entities after applying the same relation, not to mention the relation r possibly being a one-to-many relation. The triple scoring function built upon geometric distance may hardly discriminate such subtle differences in the space and thus negatively affects the quality of predictions. In principle, compared to rigid transformation in RotatE and TransE, the proposed RatE with weighted product

¹We initialize γ with 0.5 and empirically find the initialization value barely affects final performance.

Dataset	# Entity	# Relation	# Training	# Validation	# Test
WN18	40,943	18	141,442	5,000	5,000
FB15k	14,951	1,345	483,142	50,000	59,071
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466

Table 2: Statistics of four benchmark datasets of link prediction.

shares a similar inspiration with projective transformation and changes the distance between the tail entities according to spatial positions of the head entities. Consequently, besides increasing the distance between the tail entities to disambiguate entity embeddings, RatE could also decrease the distance for better support of many-to-one relations. A rigorous proof of these properties is provided in Appendix A.

3 Experiments

3.1 Experimental Setting

Dataset. We employ four widely-used link prediction benchmark, WN18, FB15K, WN18RR and FB15K-237, whose statistics are summarized in Table 2. Note, Toutanova et al., (2015) find that both WN18 and FB15K suffer from direct link problem caused by most test triples (e^1, r^1, e^2) can be found in the training or valid set with another relation, e.g., (e^1, r^2, e^2) or (e^2, r^2, e^1) .

- WN18 (Bordes et al., 2013) is extracted from WordNet (Miller, 1995), a knowledge graph composed of English phrases and lexical relations between them.
- FB15k (Bordes et al., 2013) is extracted from Freebase (Bollacker et al., 2008), which is a large-scale knowledge graph consisting real-world named entities and their relationships.
- FB15k-237 (Toutanova et al., 2015) is a subset of FB15k by 1) removing near-duplicate and inverse triples, and 2) filtering out the direct links to avoid data leakage.
- WN18RR (Dettmers et al., 2017) is a subset of WN18 following the same processes as FB15k-237.

Training Setting. The ranges of the hyper-parameters for grid search are elaborated in the following. Embedding dimension $d \in \{250, 500, 1000\}$, batch size $\in \{512, 1024, 2048\}$, and fixed margin $\lambda \in \{6, 9, 12, 18\}$. By following previous works, all entities and relation embeddings are randomly initialized under uniform distribution. The initialization range of entities is $[-\lambda/d, +\lambda/d]$ for both real and imaginary parts, and the initialization range of relations is $[0, 2\pi]$ with $|r| = 1$ in complex space. Our model is implemented using PyTorch on a single Titan V GPU. We use minibatch SGD with Adam optimizer, where the learning rate is set to 5×10^{-5} without decay.

Evaluation Metrics. Following Bordes et al. (2013), we use “*filtered*” setting to calculate evaluation metrics during test: In either head or tail entity corruption, all correct triples in train/dev/test except the current oracle test triple are removed to avoid affecting rank. Given all candidate triples ranked according to the score function $f(h, r, t)$, we use the standard evaluation metrics on link prediction tasks: 1) *mean rank* (MR) to describe the mean rank of the oracle test triples, 2) *mean reciprocal rank* (MRR), and 3) *Hits@N* ($N=1, 3, 10$) to denotes the ratio of the oracle test triples ranked in top- N .

Comparative Approach. We compare RatE with several strong graph embedding approaches, especially the trans-based approaches to which RatE belongs. In particular, for trans-based approaches, we mainly consider TransE (Bordes et al., 2013) in real space and RotatE (Sun et al., 2019) in complex space. For semantic matching approaches, we consider DistMult (Yang et al., 2015), HolE (Nickel et al., 2016), ComplEx (Trouillon et al., 2016), ConvE (Dettmers et al., 2017) and QuatE (Zhang et al., 2019). For most approaches, we copy results from the original paper or (Sun et al., 2019) except explanations.

3.2 Evaluation on Link Prediction

Link prediction results on the four datasets are shown in Table 3 and Table 4. It is observed that the proposed RatE is able to achieve new state-of-the-art results in terms of most metrics compared to previous

Method	WN18					FB15k				
	MR	MRR	Hits@10	Hits@3	Hits@1	MR	MRR	Hits@10	Hits@3	Hits@1
TransE	-	.495	.943	.888	.113	-	.463	.749	.578	.297
DistMult	655	.797	.946	-	-	42	.798	.893	-	-
HolE	-	.938	.949	.945	.930	-	.524	.739	.613	.402
ComplEx	-	.941	.947	.945	.936	-	.692	.840	.759	.599
ConvE	374	.943	.956	.946	.935	51	.657	.831	.723	.558
RotatE	309	.949	.959	.952	.944	40	.797	.884	.830	.746
QuatE	388	.949	.960	.954	.941	41	.770	.878	.821	.700
RatE	180	.950	.962	.953	.944	24	.810	.898	.859	.724

Table 3: Link prediction results on WN18 and FB15k. The results of QuatE are reported without type-constraint.

Method	WN18RR					FB15k-237				
	MR	MRR	Hits@10	Hits@3	Hits@1	MR	MRR	Hits@10	Hits@3	Hits@1
TransE	3384	.226	.501	-	-	357	.294	.465	-	-
DistMult	5110	.430	.490	.440	.390	254	.241	.419	.263	.155
ComplEx	5261	.440	.510	.460	.410	339	.247	.428	.275	.158
ConvE	4187	.430	.520	.440	.400	244	.325	.501	.356	.237
RotatE	3340	.476	.571	.492	.428	177	.338	.533	.375	.241
QuatE	3472	.481	.564	.500	.436	176	.311	.495	.342	.221
RatE	2860	.488	.590	.506	.441	172	.344	.541	.382	.261

Table 4: Link prediction results on WN18RR and FB15k-237. Values in bold denote the best results.

graph embedding approaches. Overall, compared with the baseline model RotatE, RatE merely employs several additional parameters to deliver significant improvement. To the best of our knowledge, RotatE is previous the best trans-based graph embedding approach and belongs to the same category as RatE. RatE also outperforms previous state-of-the-art semantic matching graph embedding approach, QuatE, which is defined in hypercomplex space and requires more computational overheads.

Specifically, since WN18 and FB15k suffer from the direct link problem as detailed above, it is observed that the baselines and our proposed RatE obtain comparable results in all metrics. For example, Dettmers et al. (2017) find that using a rule-based model to learn the inverse relations achieves competitive results on WN18RR. This explains why our improvement is marginal in these two datasets. Moreover, since WN18RR and FB15k-237 are presented to solve the problem in WN18 and FB15k respectively, the evaluation results on WN18RR and FB15k-237 are more canonical to measure the capability in link prediction. As shown in Table 4, the proposed RatE brings a more noticeable improvement in contrast to previous approaches.

3.3 Ablation Study

We conduct an extensive ablation study in Table 5 to verify the effectiveness of each proposed part. We first replace the relation-adaptive translation function with a shared weighted product among all relations (i.e., “RatE w/o relation-adaptive”), and observe a performance drop. And the weighted product further degenerates to standard complex product (i.e., RatE w/o weighted product), which only results in a slight drop. This suggests the proposed weighted product should be coupled with relation-adaptation to maximize its effectiveness. Then, removing L1 regularization of $\mathbf{W}^{(r)}$ in Eq.(7) and the proposed local-cognitive negative sampling leads to 0.6% and 2.6% Hits@10 drops respectively. Note “Rate w/o negative sampling” denotes using a uniform negative sampling method instead of our proposed local-

Method	MR	MRR	Hits@10	Hits@3	Hits@1
RatE full	2860	.488	.590	.506	.441
RatE w/o relation-adaptive	3278	.478	.579	.498	.432
RatE w/o weighted product	3115	.479	.576	.492	.432
RatE w/o $\mathbf{W}^{(r)}$ L1 reg	2921	.482	.584	.499	.435
RatE w/o negative sampling	3180	.471	.564	.478	.428
RatE w/o ALL	3450	.465	.556	.476	.410

Table 5: Ablation study on WN18RR.

Relation Pattern	Relation Name	$\ \mathbf{W}^{(r)}\ _1$	RatE	RotatE	TransE
Symmetry	<i>verb_group</i>	2.3	0.98	0.97	0.87
	<i>derivationally_related_form</i>	2.5	0.97	0.97	0.93
	<i>also_see</i>	2.3	0.70	0.73	0.59
Antisymmetry	<i>instance_hyponym</i>	6.1	0.56	0.54	0.22
	<i>synset_domain_topic_of</i>	3.3	0.49	0.49	0.19
	<i>member_of_domain_usage</i>	6.6	0.50	0.49	0.42
	<i>member_of_domain_region</i>	6.1	0.48	0.45	0.35
	<i>member_meronym</i>	8.7	0.54	0.38	0.04
	<i>has_part</i>	8.1	0.40	0.35	0.04
	<i>hyponym</i>	7.1	0.30	0.27	0.02
Micro Mean	-	-	0.59	0.57	0.38

Table 6: Test performance in Hits@10 regarding different relation patterns and the corresponding relations on WN18RR. $\|\mathbf{W}^{(r)}\|_1$ is used to measure the complexity of the proposed relation-adaptive translation function. Since only three triples with relation “similar_to” appear in the test set of WN18RR, we omit this relation.

Relation Type	Tail Prediction (Hits@10)				Head Prediction (Hits@10)			
	1-to-1	1-to-M	M-to-1	M-to-M	1-to-1	1-to-M	M-to-1	M-to-M
TransE (Bordes et al., 2013)	.879	.671	.964	.910	.894	.972	.567	.880
RotatE (Sun et al., 2019)	.923	.713	.961	.922	.922	.967	.602	.893
RatE*	.926	.801	.968	.924	.927	.971	.724	.895

Table 7: Performance on FB15k regarding different relation types, including injective (i.e., 1-to-1) and non-injective (e.g., 1-to-M) relations. *We replace the proposed local-cognitive negative sampling in RatE with self-adversarial one from RotatE.

cognitive negative sampling. Lastly, when removing all the proposed parts, the model is equivalent to its baseline RotatE without self-adversarial negative sampling, which results in inferior performance.

3.4 Analysis of Relation-Adaptive Translation Function

A major difference between RatE and previous trans-based graph embedding approaches (e.g., RotatE) is that a learnable relation-adaptive translation function is used in RatE to capture the translating relationship. To measure the expressive power of RatE, it is significant to investigate the learned weights in each relation-specific weighted product. As shown in Table 6, the L1 norm of learned $\mathbf{W}^{(r)}$ for symmetric relation is obviously less than that of antisymmetric relation. In particular, with the redundancy of complex number product removed, RatE preserves the ability to handle symmetric relations and achieves competitive results. For example, $\mathbf{W}^{(r)} = [[1.0, 0.1, 0.0, 0.1]; [0.0, 0.1, 1.0, 0.0]]$ is learned for relation “verb_group”. In contrast, RatE tends to construct expressively powerful translation function for anti-symmetric relations and achieves much better performance across these relations than previous models. For example, $\mathbf{W}^{(r)} = [[0.9, 1.2, 0.0, -1.7]; [0.0, 1.2, 0.9, -1.2]]$ is learned for relation “hyponym”.

3.5 Performance on Non-Injective Relations

By following Sun et al. (2019), we also evaluate the proposed RatE on different types including one injective relation type (i.e., one-to-one) and three non-injective relation types (i.e., one-to-many, many-to-one and many-to-many). As shown in Table 7, although RatE delivers similar Hits@10 values to RotatE on the injective relation type, it significantly surpasses both TransE and RotatE on the non-injective relation types. The improvements are especially substantial in 1-to-M relation (+8.8%) on tail prediction and M-to-1 (+12.2%) on head prediction, which verifies RatE’s capability in handling one-to-many relations. Coupled with the theoretical proof in §2.5, this also indirectly verifies that RatE is able to alleviate the embedding ambiguity problem posted by one-to-many relations.

3.6 Analysis of Negative Sampling

As negative sampling is crucial for a model to learn structured knowledge, we evaluate RatE with different negative sampling methods. “Local-cognitive w/o self-adv loss” can be viewed as only using prior knowledge from local closed-world assumptions (Krompaß et al., 2015). The experimental results shown in Table 8 demonstrate that compared with uniform sampling, both self-adversarial sampling and type-constraint training technique (i.e., Local-cognitive w/o self-adv loss) contribute to performance

Negative Sampling Method	WN18RR					FB15k-237				
	MR	MRR	Hits@10	@3	@1	MR	MRR	Hits@10	@3	@1
Uniform	3180	.471	.564	.478	.428	224	.320	.525	.374	.220
Self-adversarial (Sun et al., 2019)	3114	.480	.576	.481	.433	177	.339	.536	.375	.244
Local-cognitive w/o self-adv loss	3094	.479	.577	.489	.434	180	.340	.538	.374	.241
Local-cognitive (ours)	2860	.488	.590	.506	.441	172	.344	.541	.382	.261

Table 8: Performance of RatE with different negative sampling methods.

Method	WN18RR					FB15k-237					#θ
	MR	MRR	Hits@10	@3	@1	MR	MRR	Hits@10	@3	@1	
TuckER (Balazevic et al., 2019)	-	.470	.526	.482	.443	-	.358	.544	.394	.266	$d_e^2 d_r$
RatE (ours)	2860	.488	.590	.506	.441	172	.344	.541	.382	.261	$8 \mathcal{R} $

Table 9: Performance comparison between TuckER and RatE on WN18RR/FB15k-237. “#θ” denotes the number of learnable parameters only for scoring, where d_e and d_r are the embedding sizes of entity and relation respectively.

improvement. The results also emphasize the effectiveness of our proposed local-cognitive negative sampling method, a non-trivial integration of the both above, in structured knowledge learning.

3.7 Analysis of Efficiency

Lastly, we discuss RatE’s efficiency that is mainly brought by the following two factors. On the one hand, in line with previous trans-based graph embedding approaches, RatE only employs fast translation function and geometric distance measurement. On the other hand, even if a relation-adaptive translation function with weighted product is used in translating process, the function with few parameters has low time and space complexities. We compare RatE with a semantic matching graph embedding method TuckER (Balazevic et al., 2019) that uses a weight tensor to score a triple. As shown in Table 9, with competitive performance, TuckER requires much more learnable parameters than RatE for scoring. For example, TuckER has a weight tensor with 1, 200, 000 parameters on WN18RR, whereas RatE only requires 88 parameters for all the eleven relations.

4 Related Work

Unlike semantic matching graph embedding approaches (Nickel et al., 2011; Dettmers et al., 2017; Balazevic et al., 2019; Zhang et al., 2019) require additional overheads to score a triple, this work is in line with trans-based graph embedding approaches that employ an efficient translation function defined in a latent space. TransE (Bordes et al., 2013) is the most representative trans-based approach, which embeds entities/relations in real vector space and utilizes the relations as translations. It optimizes score function towards “ $\mathbf{h} + \mathbf{r} = \mathbf{t}$ ”. Several recent trans-based approaches (Wang et al., 2014; Lin et al., 2015; Ji et al., 2015; Ebisu and Ichise, 2018) can be viewed as extensions of TransE. More recently, RotatE (Sun et al., 2019), as a state-of-the-art trans-based approach, represents the entities/relations in complex vector space and formulates the translating process as a rotation in complex space. Also related to this work, many negative sampling methods (Cai and Wang, 2018; Sun et al., 2019) are proposed to effectively learn structured knowledge. KBGAN (Cai and Wang, 2018) uses knowledge graph embedding model as a negative sample generator to fool the main embedding model (i.e., the discriminator in GANs). In contrast, self-adversarial learning (Sun et al., 2019) scores a certain number of uniformly-sampled negative samples based on current model, and utilizes the scores to perform a weighted loss function. Lastly, this work is also related to using prior knowledge in graphs for training. Type-constraint method (Krompaß et al., 2015), which is based on local closed-world assumptions, corrupts heads (or tails) from relation-specific domain (or range).

5 Conclusion

In this paper, we study a novel trans-based graph embedding approach, called Relation-adaptive translating Embedding (RatE), for knowledge graph completion. It is based on the proposed relation-adaptive translation function with a novel weighted product in complex space, which not only improves represent-

ing and modeling capacities but also alleviates embedding ambiguity problem caused by non-injective relations. Therefore, RatE achieves a better trade-off between interpretability and expressive power than previous trans-based approaches. Moreover, a local-cognitive negative sampling method is also presented to seamlessly integrate prior knowledge with self-adversarial learning for effective learning. In experiments, RatE achieves state-of-the-art performance on four link prediction benchmarks. Extensive ablation study and analyses further provide comprehensive insights into RatE.

Acknowledgement

This research was funded by the Australian Government through the Australian Research Council (ARC) under the grant of LP180100654. The authors would like to appreciate anonymous reviewers for their insightful and constructive feedback.

References

- Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. 2019. Tucker: Tensor factorization for knowledge graph completion. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5184–5193. Association for Computational Linguistics.
- Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In Jason Tsong-Li Wang, editor, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250. ACM.
- Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.
- Liwei Cai and William Yang Wang. 2018. KBGAN: adversarial learning for knowledge graph embeddings. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 1470–1480. Association for Computational Linguistics.
- Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2017. Convolutional 2d knowledge graph embeddings. *CoRR*, abs/1707.01476.
- Takuma Ebisu and Ryutaro Ichise. 2018. Toruse: Knowledge graph embedding on a lie group. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1819–1826. AAAI Press.
- Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 221–231. Association for Computational Linguistics.
- Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 687–696. The Association for Computer Linguistics.
- Denis Krompaß, Stephan Baier, and Volker Tresp. 2015. Type-constrained representation learning in knowledge graphs. *CoRR*, abs/1508.02593.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In Blai Bonet and Sven Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2181–2187. AAAI Press.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Bonan Min, Ralph Grishman, Li Wan, Chang Wang, and David Gondek. 2013. Distant supervision for relation extraction with an incomplete knowledge base. In Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff, editors, *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 777–782. The Association for Computational Linguistics.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In Lise Getoor and Tobias Scheffer, editors, *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 809–816. Omnipress.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. 2016. Holographic embeddings of knowledge graphs. In Dale Schuurmans and Michael P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 1955–1961. AAAI Press.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In Aldo Gangemi, Roberto Navigli, Maria-Ester Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases. In Lluís Márquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 1499–1509. The Association for Computational Linguistics.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2071–2080. JMLR.org.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In Carla E. Brodley and Peter Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1112–1119. AAAI Press.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi, editors, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 353–362. ACM.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 2731–2741.

A Proof of Spatial Transformation

Suppose h and h' are two entities with similar embedding hold, and two triples (h, r, t) and (h', r, t') need to be distinguished.

TransE:

$$h_i + r_i = t_i \wedge h'_i + r_i = t'_i \quad (10)$$

$$\|t_i - t'_i\| = \sqrt{(h_i + r_i - h'_i - r_i)^2} = \|h_i - h'_i\|. \quad (11)$$

RotatE:

$$h_i r_i = t_i \wedge h'_i r_i = t'_i \quad (12)$$

Suppose $h_i - h'_i = a + bi, r_i = c + di$ and $\|r_i\| = \sqrt{c^2 + d^2} = 1$,

$$\|t_i - t'_i\|^2 = \|(h_i - h'_i)r_i\|^2 \quad (13)$$

$$= \|(ac - bd) + (ad + bc)i\|^2 \quad (14)$$

$$= (ac - bd)^2 + (ad + bc)^2 \quad (15)$$

$$= a^2c^2 - 2abcd + b^2d^2 + a^2d^2 + 2abcd + b^2c^2 \quad (16)$$

$$= a^2(c^2 + d^2) + b^2(c^2 + d^2) \quad (17)$$

$$= a^2 + b^2 \quad (18)$$

$$= \|h_i - h'_i\|^2 \quad (19)$$

$$\|t_i - t'_i\| = \|h_i - h'_i\|. \quad (20)$$

For RatE:

$$h_i \otimes_{r_i} r_i = t_i \wedge h'_i \otimes_{r_i} r_i = t'_i \quad (21)$$

$$t_i - t'_i = (h_i - h'_i) \otimes_{r_i} r_i \quad (22)$$

Suppose $h_i - h'_i = a + bi, r_i = c + di$ and $\|r_i\| = \sqrt{c^2 + d^2} = 1$,

$$\|t_i - t'_i\|^2 = \|(h_i - h'_i)r_i\|^2 \quad (23)$$

$$= \|(w_1ac + w_2ad + w_3bc + w_4bd) + (w_5ac + w_6ad + w_7bc + w_8bd)i\|^2 \quad (24)$$

$$= (w_1ac + w_2ad + w_3bc + w_4bd)^2 + (w_5ac + w_6ad + w_7bc + w_8bd)^2 \quad (25)$$

here we take a special case, $w = [1, 0, 0, -2, 0, 1, 1, 0]$,

$$= a^2c^2 - 4abcd + 4b^2d^2 + a^2d^2 + 2abcd + b^2c^2 \quad (26)$$

$$= a^2(c^2 + d^2) + b^2(c^2 + d^2) - 2abcd + 3b^2d^2 \quad (27)$$

$$= \|h_i - h'_i\|^2 + bd(3bd - 2ac). \quad (28)$$

Then it is quite clear that,

$$if \ bd(3bd - 2ac) = 0 \ then \ \|t_i - t'_i\| = \|h_i - h'_i\| \quad (29)$$

$$if \ bd(3bd - 2ac) > 0 \ then \ \|t_i - t'_i\| > \|h_i - h'_i\| \quad (30)$$

$$if \ bd(3bd - 2ac) < 0 \ then \ \|t_i - t'_i\| < \|h_i - h'_i\|. \quad (31)$$