OpticE: A Coherence Theory-Based Model for Link Prediction

Xiangyu Gui, Feng Zhao [*], Langjunqing Jin, Hai Jin
National Engineering Research Center for Big Data Technology and System,
Services Computing Technology and System Lab, Cluster and Grid Computing Lab
School of Computer Science and Technology, Huazhong University of Science and Technology, China
zhaof@hust.edu.cn

Abstract

Knowledge representation learning is a key step required for link prediction tasks with knowledge graphs (KGs). During the learning process, the semantics of each entity are embedded by a vector or a point in a feature space. The distance between these points is a measure of semantic similarity. However, in a KG, while two entities may have similar semantics in some relations, they have different semantics in others. It is ambiguous to assign a fixed distance to depict the variant semantic similarity of entities. To alleviate the semantic ambiguity in KGs, we design a new embedding approach named OpticE, which is derived from the well-known physical phenomenon of optical interference. It is a lightweight and relation-adaptive model based on coherence theory, in which each entity's semantics vary automatically regarding different relations. In addition, a unique negative sampling method is proposed to combine the multimapping properties and self-adversarial learning during the training process. The experimental results obtained on practical KG benchmarks show that the OpticE model, with elegant structures, can compete with existing link prediction methods.

1 Introduction

Knowledge graphs (KGs) consist of sets of triplets that can represent real-world concepts, common sense information or facts. Each triplet (h, r, t) indicates a directional relation r from the head entity h to the tail entity t. WordNet (Miller, 1995), Freebase (Bollacker et al., 2008) and Wikidata (Vrandečić and Krötzsch, 2014) are the most well-known KGs with tremendous numbers of entities and relations. These KGs play important roles in a range of areas, such as natural language processing tasks (Cao et al., 2021), question answering applications (Bosselut et al., 2021) and recommendation systems (Shao et al., 2021). Research on KGs is thriving in both academic and industrial communities.

Link prediction is a fundamental problem when addressing KGs. To learn the hidden patterns from the observed triplets, extensive investigations have been performed to embed entities and relations into a continuous semantic space; this is known as KG embedding learning (Rossi et al., 2021). Among these embedding models, from a mathematical perspective, TransE (Bordes et al., 2013), TransH (Wang et al., 2014) and TransR (Lin et al., 2015) are translation-based models that use translation transformations to characterize the existence of triplets. Other methods (e.g., RESCAL (Nickel et al., 2011) and DistMult (Yang et al., 2015)) tackle this problem with matrix or tensor multiplication. Recently, the embedding problem has been discussed in the complex domain (e.g., ComplEx (Trouillon et al., 2016), RotatE (Sun et al., 2019) and HAKE (Zhang et al., 2020)). These methods have achieved great performance on related tasks. While our models are discussed in the complex domain, they are designed under an interference framework.

The semantic ambiguity caused by multimapping relations was ignored in most previous methods. As illustrated in Figure 1 (a), the semantics of *New York* and *Washington D.C.* are similar under the relation *Located_in*, while they are quite different under the relation *Capital_of* because *New York* is not a capital. Previous methods such as TransE and RotatE suffer from this ambiguity problem. As shown in Figure 1 (b), it is difficult to disambiguate *New York* from *Washington D.C.* for the relation *Capital_of* in RotatE. To alleviate this problem, TransH utilizes a hyperplane to project entities into independent semantic space. RatE (Huang et al., 2020) integrates a semantic matching method with an extra relation-adaptive matrix. Instead of relying on extra settings, we tackle this problem by the intrinsic attribute of the

[*] Corresponding author

Proceedings of the 29th International Conference on Computational Linguistics, pages 4892–4901
October 12–17, 2022.

rotational semantic space.

We are motivated by the optical interference phenomenon (bright or dark fringes appear after the superposition of two beams of light), because (i) it is intuitive to simulate positive cases and negative cases utilizing distinct dark and bright fringes; and (ii) since coherence theory in the complex field is sufficiently sophisticated, it can play a theoretical guiding role. As OpticE is one type of rotation-based model, it can be illustrated with the rotation transformation shown in Figure 1 (c). The entities *New York* and *Washington D.C.* are mapped into different relation-adaptive orbits with the same phase difference. The two entities are close to each other in the orbit of *Located_in*, while in the orbit of relation *Capital_of* they no longer overlap. By embedding entities into orbits with different semantic densities, OpticE can take control of the similarity of entities regarding specific relations. Generally, multimapping relations map entities into small-radius orbits in dense semantic space, while one-to-one relations rely on large-radius orbits in sparse semantic space to distinguish entities. This will be verified in our experiments.

To optimize OpticE effectively, we optimize self-adversarial sampling (Sun et al., 2019) by integrating a Bernoulli filtering process. The proposed method can leverage the high efficiency of self-adversarial sampling and reduce the performance loss caused by false negative labels. Tested on WN18, FB15k, WN18RR and FB15k-237, the widely recognized benchmark datasets for link prediction, the lightweight OpticE model is comparable to the previous rotation-based models. OpticE's ability to adapt to multimapping relations is also verified in our experiments. Our contributions can be summarized as follows:

- **Coherence theory** is introduced into the KG reasoning task for the first time to make the embeddings of KGs in a complex field more intuitive and understandable.
- **A relation-adaptive amplitude modulation technique**, which is simple, effective and lightweight, is developed to alleviate the multimapping problem in KGs.
- **A novel negative sampling method** is provided to train our model precisely and effectively; this approach can be generally applied to other models.

2 Related Work

All the typical related works are listed in Table 1 with their model realizations and complexities in terms of the number of utilized parameters. They are divided into two groups according to whether they are calculated in a real or complex number field.

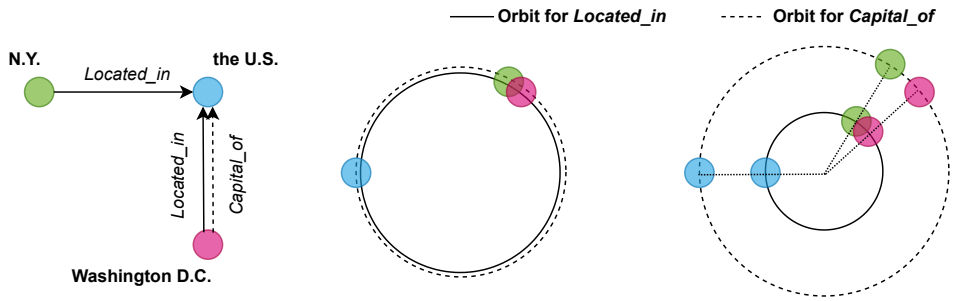
Real number field models. The most important and influential model is TransE (Bordes et al., 2013), which takes relations as translation transformation by simply assuming that $h + r = t$. However, its expressiveness is limited, especially when representing a multimapping relation. TransH (Wang et al., 2014) exploits an extra projection transformation and manages to eliminate the ambiguity in TransE. DistMult (Yang et al., 2015) uses a semantic matching method while ignoring the directionality of the relations.

Complex field models. ComplEx (Trouillon et al., 2016) utilizes complex numbers in KG embeddings and achieves great performance. It can be considered a complex version of DistMult, but it overcomes the shortcomings of DistMult via tail conjugation. RotatE (Sun et al., 2019) surpasses all the previous models by rotation transformation; similar to TransE, the model assumes that $h \circ r = t$ and suffers from the same ambiguity problem. Based on RotatE, RatE (Huang et al., 2020) defines a relation-specific weight W_r with an extra 8 parameters per relation to solve the ambiguity problem. HAKE (Zhang et al., 2020) employs a hierarchical structure simulated by a modulus, but the model is complicated and difficult to train.

OpticE (ours) belongs to the category of complex field models and provides a novel way to solve the semantic ambiguity problem in RotatE. Like RatE and TransH, it has a relation-adaptive strategy that can retrieve each entity's different semantics. However, our strategy is much simpler due to the use of a relation-modulated amplitude. In the aspect of negative sampling, we adopt multiple negative sampling (Trouillon et al., 2016) and self-adversarial strategy (Sun et al., 2019) in our model, but optimize the process with a Bernoulli distribution (Wang et al., 2014) according to the multimapping relations.

3 The Coherence Theory-Based Model

In this part, we briefly introduce coherence theory in §3.1. In §3.2, we provide three ways to modulate an interference model and demonstrate the ad-



(a) KG (b) RotatE (c) OpticE

Figure 1: Illustration of OpticE. (a) A toy example is given. (b) RotatE is unable to disambiguate *New York* from *Washington D.C.* for the relation *Capital_of*. (c) The entities *New York* and *Washington D.C.* are mapped into different relation-adaptive orbits with the same phase difference.

Table 1: KG embedding models and their storage complexities of parameters, where $\circ$ denotes the Hadamard product, $||\cdot||_{1/2}$ denotes the L1 or L2 norm, and $\odot$ denotes a specific weighted product defined in RatE. k is the dimensionality. N_e and N_r are the numbers of entities and relations.

Model	Score function	Parameters	Complexity
TransE	$-\|\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|\|_{1/2}$	$\mathbf{h},\mathbf{r},\mathbf{t}\in\mathbb{R}^k$	$\mathcal{O}(N_e k+N_r k)$
TransH	$-\|\|(\mathbf{h}-\mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r)+\mathbf{d}_r-(\mathbf{t}-\mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r)\|\|_2^2$	$\mathbf{h},\mathbf{t},\mathbf{d}_r,\mathbf{w}_r\in\mathbb{R}^k$	$\mathcal{O}(N_e k+2N_r k)$
DistMult	$\mathbf{h}^\top diag(\mathbf{r})\mathbf{t}$	$\mathbf{h},\mathbf{r},\mathbf{t}\in\mathbb{R}^k$	$\mathcal{O}(N_e k+N_r k)$
ComplEx	$Re(\mathbf{h}^\top diag(\mathbf{r})\bar{\mathbf{t}})$	$\mathbf{h},\mathbf{r},\mathbf{t}\in\mathbb{C}^k$	$\mathcal{O}(2N_e k+2N_r k)$
RotatE	$-\|\|\mathbf{h}\circ\mathbf{r}-\mathbf{t}\|\|_1$	$\mathbf{h},\mathbf{r},\mathbf{t}\in\mathbb{C}^k, \|r_i\|=1$	$\mathcal{O}(2N_e k+N_r k)$
RatE	$-\|\|\mathbf{h}\odot_{\mathbf{W}(r)}\circ\mathbf{r}-\mathbf{t}\|\|_1$	$\mathbf{h},\mathbf{r},\mathbf{t}\in\mathbb{C}^k, \mathbf{W}\in\mathbb{R}^{2\times 4}$	$\mathcal{O}(2N_e k+N_r(k+8))$
HAKE	$-\|\|\mathbf{h}_m\circ\mathbf{r}_m-\mathbf{t}_m\|\|_2-$ $\lambda\|\|\sin((\mathbf{h}_p+\mathbf{r}_p-\mathbf{t}_p)/2)\|\|_1$	$\mathbf{h}_m,\mathbf{r}_m,\mathbf{t}_m,\mathbf{h}_p,\mathbf{r}_p,\mathbf{t}_p\in\mathbb{R}^k$	$\mathcal{O}(2N_e k+2N_r k)$
OpticE (ours)	$-\|\|\langle\mathbf{M}_\mathbf{r}(\mathbf{h}),\mathbf{M}'_\mathbf{r}(\mathbf{t})\rangle\|\|_1$	$\mathbf{h}$ and $\mathbf{t}$ are light sources, $\mathbf{M}(\cdot)$ is modulator	$\mathcal{O}(N_e k+3N_r k)$

vantage of the proposed OpticE. In §3.3, the model training process is introduced with the proposed Bernoulli self-adversarial sampling method.

3.1 Interference under Coherence Theory

Optical interference is a common phenomenon in physics that can be interpreted formally by the superposition of waves in coherence theory. We briefly introduce this theory. More details can be found in (Hecht, 2016).

Monochromatic harmonic light can be represented as a point in the complex field as

$$E = A\exp[i\varphi], \tag{1}$$

where φ is the phase and A is the amplitude (or modulus). The intensity of the light synthesized by E_1 and E_2 can be defined as

$$\begin{aligned} I &= \langle E_1, E_2\rangle \\ &= A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_1-\varphi_2). \end{aligned} \tag{2}$$

If and only if $A_1 = A_2$ and phase difference is π, the intensity tends to be 0, which is called **total destructive interference**. In our models, we utilize this state to indicate positive triplets in KGs.

3.2 The Proposed Models

Based on the coherence theory, we take the entities as a series of light sources and the relations as a type of modulation performed on their phases and the amplitudes. Then, the existence of triplets can be indicated by the intensity of the lights after superposition.

To be more intuitive, we modify Equation (2) as

$$I = A_1^2 + A_2^2 - 2A_1A_2\cos(\varphi_1-\varphi_2). \tag{3}$$

Then, the **total destructive interference** condition becomes $A_1 = A_2$ and $\varphi_1 = \varphi_2$. Given a triplet $(h,r,t)\in\mathcal{G}$, each entity consists of k (k is the dimension) light sources. By extending Equation (1), the head entity can be represented as $\mathbf{h} = [E_h^1, E_h^2, \cdots, E_h^k]$, and the tail entity is represented as $\mathbf{t} = [E_t^1, E_t^2, \cdots, E_t^k]$. The score function is

$$f_r(h,t) = -||\langle\mathbf{M}_\mathbf{r}(\mathbf{h}),\mathbf{M}'_\mathbf{r}(\mathbf{t})\rangle||_1,$$

where $\mathbf{M}_\mathbf{r}(\cdot)$ is the modulation operates on the head or the tail, and $\langle\mathbf{M}_\mathbf{r}(\mathbf{h}),\mathbf{M}'_\mathbf{r}(\mathbf{t})\rangle$ is a list of

Algorithm 1 Bernoulli self-adversarial sampling

Require: A KG $\mathcal{G}$, the given true triplet $(h, r, t) \in \mathcal{G}$, the number of negative samples n. For each relation in KG, we first get the statistics about hpt and tph (hpt is the average number of head entities per tail entity; tph is the average number of tail entities per head entity).

Ensure: The negative sample set $\mathcal{S}$ and $p(h'_i, r, t'_i)$.

$p = \frac{hpt}{hpt+tph}$ // probability to corrupt the tail
$\epsilon \leftarrow U(0, 1)$ // ϵ uniformly generated from $(0, 1]$
if $\epsilon < p$ **then**
  while $|\mathcal{S}| < n$ **do**
    $\mathcal{S} \leftarrow \mathcal{S} \cup (h, r, t')$ // to corrupt the tail
  end while
else
  while $|\mathcal{S}| < n$ **do**
    $\mathcal{S} \leftarrow \mathcal{S} \cup (h', r, t)$ // to corrupt the head
  end while
end if
for the i-th sample $(h'_i, r, t'_i) \in \mathcal{S}$, its weight $p(h'_i, r, t'_i)$ can be obtained from Equation (4).

the synthesized intensities of each dimension calculated by Equation (3). The larger the score function value is, the more likely (h, t, r) is the positive case. When the **total destructive interference** condition is met, the score reaches its maximum 0.

In this paper, three ways are designed to modulate the entities. They are summarized as **pOpticE, aOpticE** and **OpticE** models.

(i) **pOpticE with only phase modulation.** The amplitudes of eahc light sources are fixed at 1. A relation only modulates the corresponding phase. Specifically, we have

$$\mathbf{M_r}(\mathbf{h}) = \exp\left[i(\varphi_\mathbf{h} + \varphi^\mathbf{h}_\mathbf{r})\right]$$
$$\mathbf{M'_r}(\mathbf{t}) = \exp\left[i(\varphi_\mathbf{t} + \varphi^\mathbf{t}_\mathbf{r})\right].$$

The modulator provides a phase delay $\varphi^\mathbf{h}_\mathbf{r}$ to the head and $\varphi^\mathbf{t}_\mathbf{r}$ to the tail; then, the score function is

$$f_r(h, t) = ||2 - 2\cos(\varphi_\mathbf{h} + \varphi_\mathbf{r} - \varphi_\mathbf{t})||_1,$$

in which $\varphi_\mathbf{r} = \varphi^\mathbf{h}_\mathbf{r} - \varphi^\mathbf{t}_\mathbf{r}$ is the phase difference added to the head's phase by relation *r*. Then, the phase modulation process can be simplified as

$$\mathbf{M_r}(\mathbf{h}) = \exp\left[i(\varphi_\mathbf{h} + \varphi_\mathbf{r})\right]$$
$$\mathbf{M'_r}(\mathbf{t}) = \exp\left[i(\varphi_\mathbf{t})\right].$$

Similar to other rotation-based models, our models can infer **symmetric, anti-symmetric, inversion** and **composition** relations because all these models rely on the modulated phase difference $(\varphi_\mathbf{h} + \varphi_\mathbf{r} - \varphi_\mathbf{t})$.

(ii) **aOpticE with additional entity-adaptive amplitudes.** The amplitudes of the sources are determined by the specific entities themselves. Then, we obtain

$$\mathbf{M_r}(\mathbf{h}) = \mathbf{A_h} \exp\left[i(\varphi_\mathbf{h} + \varphi_\mathbf{r})\right]$$
$$\mathbf{M'_r}(\mathbf{t}) = \mathbf{A_t} \exp\left[i(\varphi_\mathbf{t})\right].$$

Similar to RotatE, the amplitudes of the entities in aOpticE remain unalterable under different relations.

(iii) **OpticE with relation-adaptive amplitude modulation.** The amplitudes of the sources are determined by the entities and relations, simultaneously. We have

$$\mathbf{M_r}(\mathbf{h}) = \mathbf{A_{r,h}} \exp\left[i(\varphi_\mathbf{h} + \varphi_\mathbf{r})\right]$$
$$\mathbf{M'_r}(\mathbf{t}) = \mathbf{A_{r,t}} \exp\left[i(\varphi_\mathbf{t})\right],$$

in which $\mathbf{A_{r,h}}$ and $\mathbf{A_{r,t}}$ are the amplitudes of the entities modulated by the relations. In this paper, we provide a specific format of $\mathbf{A_{r,h}}$ and $\mathbf{A_{r,t}}$ to map the entities into relation-adaptive rotational orbits with different semantic density:

$$\mathbf{A_{r,h}} = 1 + \lambda\cos(\varphi_\mathbf{h} + \phi_\mathbf{r})$$
$$\mathbf{A_{r,t}} = 1 + \lambda\cos(\varphi_\mathbf{t} + \phi'_\mathbf{r}),$$

where $\lambda \in (0, 1)$ is the coefficient. For OpticE, the amplitudes, which fluctuate between $[1-\lambda, 1+\lambda]$, are controlled by different relations through the extra parameters $\phi_\mathbf{r}$ and $\phi'_\mathbf{r}$.

OpticE is a lightweight model with high storage efficiency, as demonstrated in Table 1. By reusing phase parameters $\varphi_\mathbf{h}$ and $\varphi_\mathbf{t}$ in $\mathbf{A_{r,h}}$ and $\mathbf{A_{r,t}}$, only $2N_r * k$ (N_r is the number of relations) extra parameters are needed. However, for other rotation-based models like aOpticE, we need $N_e * k$ (N_e is the number of entities) parameters to represent the amplitudes. The number of parameters is reduced by $(N_e - 2N_r) * k$, where $N_r \ll N_e$ in KGs.

3.3 Training

Negative sampling is essential to train the models well. This approach is an effective method derived from word embedding (Mikolov et al., 2013) and has been proven powerful for KG embedding

Table 2: Statistics of the datasets. The last column is the number of the observed triplets in each test, validation and training set.

Dataset	$\mathbf{N_e}$	$\mathbf{N_r}$	# Te./Val./Tr.
FB15k	14,951	1,345	59k / 50k / 483k
WN18	40,943	18	5k / 5k / 141k
FB15k-237	14,541	237	20k / 18k / 272k
WN18RR	40,943	11	3k / 3k / 87k

(Huang et al., 2020; Krompaß et al., 2015; Sun et al., 2019). Similar to (Sun et al., 2019), the self-adversarial sampling method is used:

$$\mathcal{L} = -\log\sigma(\gamma + f_r(\mathbf{h},\mathbf{t})) - \sum_{i=1}^{n} p(h'_i, r, t'_i)\log\sigma(-f_r(\mathbf{h}'_i,\mathbf{t}'_i) - \gamma),$$

where (h'_i, r, t'_i) is the i-th negative sample of the true triplet (h, r, t), γ is a fixed margin and σ is the sigmoid function. In OpticE and pOpticE, $f_r(\mathbf{h},\mathbf{t})$ need to be reduced by a proper times to fit the value of γ. Furthermore,

$$p(h'_i, r, t'_i) = \frac{\exp \alpha f_r(\mathbf{h}'_i,\mathbf{t}'_i)}{\sum_j \exp \alpha f_r(\mathbf{h}'_j,\mathbf{t}'_j)} \quad (4)$$

is the weight of the i-th negative sample, where α is the adversarial temperature. According to Equation (4), negative samples with higher scores are assigned with higher weights during the training phase. This is an effective way to filter out high-scoring negative samples and lower their scores efficiently. Apparently, this technique is better than assigning the same average weight to each sample.

However, there are obvious defects in the self-adversarial sampling methods above. During the sampling process, those missing true triplets to be predicted will inevitably suffer from a severe penalty if they are sampled, which will undermine the prediction accuracy. Here, we take advantage of the Bernoulli sampling strategy described in (Wang et al., 2014) to reduce the chance of sampling false negative samples. Instead of replacing the head and tail uniformly, we use a Bernoulli distribution to decide whether to corrupt the head or the tail. For example, given a triplet *(Biden, nationality, the U.S.)*, since *nationality* is a many-to-one relation, it is safer to change the entity *the U.S.* to obtain a negative sample than altering entity *Biden*. In Algorithm 1, the process of the Bernoulli self-adversarial sampling method is clarified.

4 Experiments

In this part, the experimental conditions are elaborated, and the performance of the proposed models is compared. Different negative sampling methods are tested. Finally, we examine the effects of the amplitude of OpticE. All the experiments are conducted on an NVIDIA-v100 GPU. The code for this paper is available on https://github.com/guixiangyu1/OpticE.

4.1 Experimental Settings

Four widely used datasets,FB15k (Bordes et al., 2013), WN18 (Bordes et al., 2013), FB15k-237 (Toutanova and Chen, 2015) and WN18RR (Dettmers et al., 2018), are benchmark datasets in our experiments. All these datasets contain plenty of multimapping instances. The statistics of these datasets are presented in Table 2. The models are trained on the training set and tested on the test set.

- **FB15k** contains tremendous entities and multimapping relations from real-world settings and is extracted from Freebase (Bollacker et al., 2008).
- **WN18** is extracted from WordNet (Miller, 1995) and is designed to manipulate the semantic relations between words and phrases.
- **FB15k-237** is a subset of FB15k. To avoid the data leakage problem in FB15k, duplicate triplets and direct links are removed in FB15k-237.
- **WN18RR** is a subset of WN18. WN18RR follows the same processes as FB15k-237.

Hyperparameters. During the training process, Adam (Kingma and Ba, 2014) is used to optimize OpticE. Grid search is used and the range of hyperparameters is set as follows: the batch size $b \in [512, 1024, 2048]$, the number of embedding dimensions $k \in [500, 1000, 1500, 2000]$, the temperature for Bernoulli self-adversarial sampling $\alpha \in [0.5, 1.0]$, the negative sample size $n \in [50, 100, 200, 256]$ and the margin $\gamma \in [6, 9, 12, 15]$. All amplitudes and phases are initialized uniformly. The range of the phases is $[0, 2\pi)$.

Evaluation. All the models are tested their link prediction performance with three standard evaluation metrics: the mean reciprocal rank (MRR), mean rank (MR) and hits at N (Hits@N). Given a true test triplet, its head entity or tail entity is

Table 3: Link prediction results on FB15k and WN18. We take TransE results from (Nickel et al., 2016), and DistMult results from (Kadlec et al., 2017). Other results are extracted as benchmarks from the corresponding original papers.

Method	FB15k					WN18				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
TransE	-	..463	.297	.578	.749	-	.495	.113	.888	.943
DistMult	42	.798	-	-	.893	655	.797	-	-	.946
HolE	-	.524	.402	.613	.739	-	.938	.930	.945	.949
ComplEx	-	.692	.599	.759	.840	-	.941	.936	.945	.947
ConvE	51	.657	.558	.723	.831	374	.943	.935	.946	.956
RotatE	40	.797	.746	.830	.884	309	.949	.944	.952	.959
RatE	**24**	**.810**	.724	**.859**	**.898**	**180**	.950	.944	.953	**.962**
pOpticE	42	.792	.742	.823	.881	290	.948	.940	.952	.960
aOpticE	45	.788	.736	.822	.879	362	.946	.940	.952	.958
OpticE	39	.804	**.756**	.837	.889	261	**.951**	**.946**	**.955**	.961

Table 4: Link prediction results on FB15k-237 and WN18RR. The results of [♠] are taken from (Dettmers et al., 2018). Other results are extracted as benchmarks from the corresponding original papers.

Method	FB15k-237					WN18RR				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
TransE	357	.294	-	-	.465	3384	.226	-	-	.501
DistMult [♠]	254	.241	.155	.263	.419	5110	.43	.39	.44	.49
ComplEx [♠]	339	.247	.158	.275	.428	5261	.44	.41	.46	.51
ConvE [♠]	244	.325	.237	.356	.501	4187	.43	.40	.44	.52
RotatE	177	.338	.241	.375	.533	3340	.476	.428	.492	.571
HAKE	-	.346	.250	.381	.542	-	**.497**	.452	**.516**	.582
RatE	172	.344	.261	.382	.541	2860	.488	.441	.506	**.590**
pOpticE	183	.329	.232	.366	.525	3182	.477	.435	.491	.559
aOpticE	186	.340	.244	.377	.536	3510	.473	.425	.489	.563
OpticE	**151**	**.359**	**.264**	**.398**	**.550**	**1930**	**.497**	**.453**	.512	.585

corrupted and we rank all the candidate triplets by their scores from $f_r(h', t)$ or $f_r(h, t')$. These ranks are calculated by filtering out all the correct triplets except the one to be predicted. Higher MRR, Hits@N and lower MR suggest better performance.

4.2 Main Results

The existing state-of-the-art models are compared with ours in terms of link prediction. They are TransE (Bordes et al., 2013), DistMult (Yang et al., 2015), ComplEx (Trouillon et al., 2016), HolE (Nickel et al., 2016), ConvE (Dettmers et al., 2018), RotatE (Sun et al., 2019), HAKE (Zhang et al., 2020) and RatE (Huang et al., 2020).

In Tables 3 and 4, the main results obtained on the link prediction task are listed. RotatE surpasses all the previous models by simultaneously modeling different kinds of relations in KGs, including symmetry, antisymmetry, inversion and composition (Sun et al., 2019). HAKE, RatE and our OpticE can be considered members of the RotatE family because they all employ the rotation transformation approach and are capable of representing the four types of relations described above. By optimizing the hierarchy and ambiguity aspects, HAKE and RatE outperform RotatE to a large extent. According to the results, by mapping entities into different relation-adaptive semantic spaces, OpticE is comparable to these two models and even exceeds them in terms of most evaluation metrics.

The performances of different modulation styles are also compared. pOpticE (with only phase modulation) can complete the link prediction task well to some extent. By assigning trainable amplitudes decided by entities, aOpticE optimizes the corresponding outcomes. OpticE, with relation-adaptive modulated amplitude, greatly surpasses aOpticE in four datasets, which verifies that it is beneficial to use relation-adaptive amplitude modulation rather than entity-adaptive amplitude modulation.

4.3 Performance for Different Relation Types

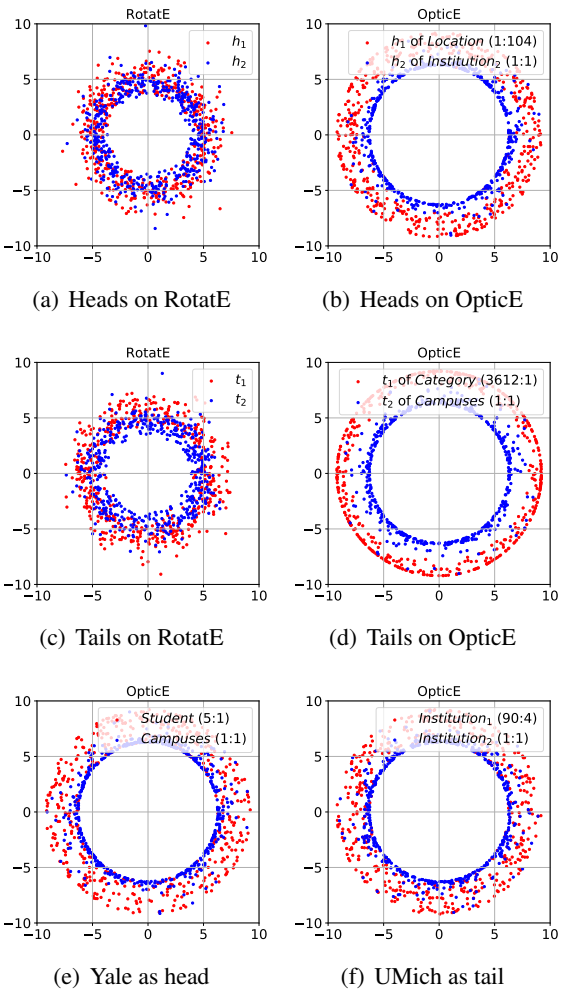
We further investigate the performance of OpticE when processing prediction tasks with different relation types. According to the general classification method given by (Wang et al., 2014), the relations are classified into one-to-one, one-to-many,

Table 5: Performance on FB15k-237 for different relation types. TransE and RotatE are trained according to the code released by (Sun et al., 2019) with their best configurations.

	Head Prediction (MRR)				Tail Prediction (MRR)			
Relation Type	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	.491	.453	.085	.255	.481	.074	.741	.361
RotatE	.501	.472	.092	.261	.488	**.075**	.748	.368
OpticE (ours)	**.505**	**.484**	**.113**	**.285**	**.499**	.073	**.777**	**.390**

Table 6: Performance of RotatE and OpticE with different negative sampling methods. For FB15k-237, the negative sample size is set to 256. For WN18RR, the negative sample size is set to 100. Other conditions remain the same.

	FB15k-237					WN18RR				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
RotatE(uni.)	186	.295	.203	.326	.479	**3226**	.471	.425	.487	.564
RotatE(adv.)	**178**	.335	.238	.374	**.531**	3338	.477	.429	.493	.574
RotatE(badv.)	179	**.337**	**.241**	**.374**	**.531**	3267	**.478**	**.430**	**.496**	**.578**
OpticE(uni.)	152	.348	.254	.385	.537	2267	.486	.442	.501	.568
OpticE(adv.)	168	.354	.260	.390	.541	2890	.488	.445	.504	.570
OpticE(badv.)	**151**	**.359**	**.264**	**.398**	**.550**	1930	**.497**	**.453**	**.512**	**.585**



(a) Heads on RotatE (b) Heads on OpticE

(c) Tails on RotatE (d) Tails on OpticE

(e) Yale as head (f) UMich as tail

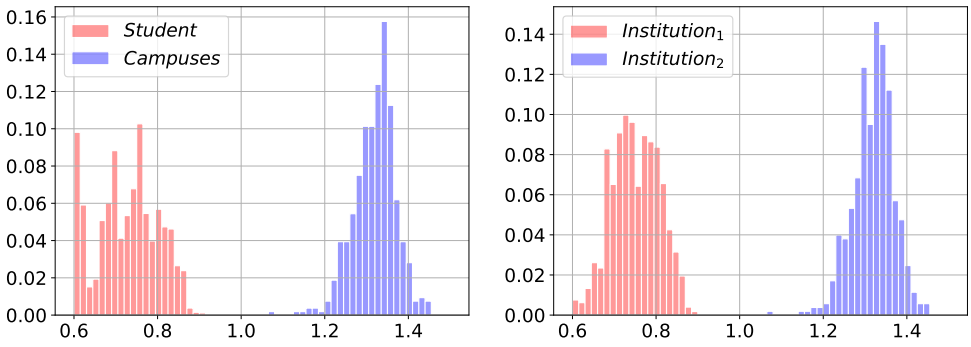
Figure 2: Visualization of several entities on FB15k-237. The Mid of h_1 and h_2 are /m/04ztj and /m/07tg4; the Mid of t_1 and t_2 are /m/08mbj5d and /m/0bwfn.

many-to-one and many-to-many types. The results obtained by OpticE on the FB15k-237 dataset are given in Table 5 and compared with those of TransE and RotatE.

We can see that OpticE achieves improvements for almost all types of relations, except the one-to-many tail prediction relations for which all the methods obtain low MRR scores. It is difficult to predict the 'many' parts of one-to-many and many-to-one relations. This is in line with the common sense notion that 'many' indicates more uncertainty. By employing relation-adaptive amplitudes, OpticE alleviates this problem to some extent.

4.4 Analysis of Negative Sampling Methods

We apply different negative sampling methods to OpticE and RotatE and record the results in Table 6. Uniform sampling (uni.), self-adversarial sampling (adv.) and Bernoulli self-adversarial sampling (badv.) are alternated in the models during the test. These methods are evaluated based on the results in Table 6. During the sampling phase, Bernoulli sampling can decrease the chance of sampling false negative samples. As indicated in the table, compared with adv. and uni. sampling, models with badv. sampling achieve the best result. From the perspective of models, OpticE outperforms RotatE with the same negative sampling method, which proves the excellent representation ability of OpticE.



(a) *Student* vs. *Campuses* (b) *Institution*$_1$ vs. *Institution*$_2$

Figure 3: Distribution histogram of the amplitudes. (a) For *Student* (many-to-one) and *Campuses* (one-to-one), we take the average amplitude of their heads and calculate the corresponding relative frequency. (b) For $Institution_1$ (many-to-many) and $Institution_2$ (one-to-one), we take the average amplitude of their tails and calculate the corresponding relative frequency.

4.5 Analysis Regarding the Amplitude

To verify the effects of the amplitude, we monitor and visualize it in OpticE. we visualize the heads and tails on FB15k-237 with several relations, as shown in Figure 2. In Figure 3, we test the amplitude distribution of different kinds of relations.

Visualization of relation-adaptive amplitudes in OpticE. In Figure 2, all the entities are mapped into a 2D space with the data of their first 500 dimensions, based on a complex coordinate system. To make the property of the amplitude more obvious, we take the logarithm operation on the amplitudes. Since the amplitudes and moduli are less than 1, the smaller radii in the figure will actually indicate larger amplitudes or moduli. The numbers in the parentheses are $(hpt : tph)$ of the relations.

In Figure 2(a) and (b), (c) and (d), RotatE mixes entities in the same space while OpticE can distinguish them by different relations. The radii under one-to-one relations ($Institution_2$ and $Campuses$) are smaller in OpticE, which means the entities are mapped into sparse semantic space with larger amplitudes. In Figure 2(e) and Figure 2(f), although *Yale* is the head entity of both relations, entity *Yale* is mapped into sparse semantic space by relation *Campuses* (one-to-one) and dense semantic space by relation *Student* (many-to-one). A similar situation occurs with *UMich* on $Institution_1$ and $Institution_2$. These visualizations prove that OpticE can tackle the semantic ambiguity problem with relation-adaptive amplitudes.

Distribution of amplitudes of different relations. The relation between the values of amplitudes and the multimapping properties of the relation is investigated. As illustrated in Figure 3, the amplitudes of injective (one-to-one) relations ($Campuses$ and $Institution_2$) are larger than those of noninjective relations ($Student$ and $Institution_1$). The reason is that for multimapping relations, the semantics of entities overlap with each other. Then with smaller amplitudes in a dense semantic space, similar semantics are more convenient to express. For the one-to-one relations, entities are exclusive and there are semantic gaps between them. With larger amplitudes in a sparse semantic space, each entity can disambiguate with each other more easily. These attributes are also demonstrated in Figure 2, showing that multimapping relations have smaller radii (red points) and one-to-one relations with larger ones (blue points).

5 Conclusion

We embed KGs with a novel optic interference perspective to tackle the link prediction problem. Guided by coherence theory, we explore three kinds of modulation methods and analyze their properties theoretically. With our proposed negative sampling method, OpticE achieves the best score in comparison with other translation- and rotation-based models. Its ability to disambiguate entities in multiple mapping relations is verified experimentally. The results suggest that we can alleviate semantic ambiguity in rotation-based models by mapping entities into relation-adaptive orbits with different semantic densities.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China under Grants No.62072203.

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26:2787–2795.

Antoine Bosselut, Ronan Le Bras, and Yejin Choi. 2021. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*.

Boxi Cao, Hongyu Lin, Xianpei Han, Le Sun, Lingyong Yan, Meng Liao, Tong Xue, and Jin Xu. 2021. Knowledgeable or educated guess? revisiting language models as knowledge bases. *arXiv preprint arXiv:2106.09231*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Thirty-second AAAI conference on artificial intelligence*.

E. Hecht. 2016. *Optics, Global Edition*. Pearson Education Limited.

Hao Huang, Guodong Long, Tao Shen, Jing Jiang, and Chengqi Zhang. 2020. Rate: Relation-adaptive translating embedding for knowledge graph completion. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 556–567.

Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. 2017. Knowledge base completion: Baselines strike back. *arXiv preprint arXiv:1705.10744*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Denis Krompaß, Stephan Baier, and Volker Tresp. 2015. Type-constrained representation learning in knowledge graphs. In *International semantic web conference*, pages 640–655. Springer.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2181–2187.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

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

Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pages 809–816.

Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. 2021. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(2):1–49.

Bilin Shao, Xiaojun Li, and Genqing Bian. 2021. A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Systems with Applications*, 165:113764.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*.

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.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, pages 1112–1119. Citeseer.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning hierarchy-aware knowledge graph embeddings for link prediction. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 3065–3072. AAAI Press.

A More Details about Coherence Theory

Given $E_1 = A_1 e^{i\varphi_1}$ and $E_2 = A_2 e^{i\varphi_2}$, after the superposition of the two light waves, the formed synthetic light is

$$\begin{aligned} E =& A_1 e^{i\varphi_1} + A_2 e^{i\varphi_2} \\ =& A_1 \cos\varphi_1 + A_2 \cos\varphi_2 \\ &+ i(A_1 \sin\varphi_1 + A_2 \sin\varphi_2). \end{aligned}$$

Then, we have

$$\begin{aligned} ||E||^2 =& (A_1 \cos\varphi_1 + A_2 \cos\varphi_2)^2 \\ &+ (A_1 \sin\varphi_1 + A_2 \sin\varphi_2)^2 \\ =& A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_1 - \varphi_2). \end{aligned}$$

The final intensity is $I = \frac{1}{2}||E||^2$. For convenience, the coefficient $\frac{1}{2}$ is ignored. The synthetic light intensity of two coherent lights can be noted as

$$\begin{aligned} I &= \langle E_1, E_2 \rangle \\ &= A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_1 - \varphi_2). \end{aligned}$$

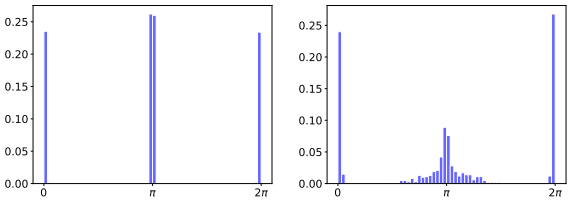
To be more intuitive, we modify the intensity as

$$\begin{aligned} I &= ||E_1 - E_2||^2 \\ &= A_1^2 + A_2^2 - 2A_1A_2\cos(\varphi_1 - \varphi_2). \end{aligned}$$

Then, the phase difference between φ_1 and φ_2 is important during the process ($k \in \mathcal{Z}$):

$$I = \begin{cases} (A_1 + A_2)^2, & \varphi_1 - \varphi_2 = (2k+1)\pi, \\ A_1^2 + A_2^2, & \varphi_1 - \varphi_2 = (\frac{1}{2} + k)\pi, \\ (A_1 - A_2)^2, & \varphi_1 - \varphi_2 = 2k\pi. \end{cases}$$

When $\cos(\varphi_1 - \varphi_2) = 1$, the intensity reaches its maximum as $(A_1 + A_2)^2$. When $\cos(\varphi_1 - \varphi_2) = 0$, the interference effect disappears, and intensity is the sum of the individual intensities as $A_1^2 + A_2^2$. The extreme case is **total destructive interference**, in which the waves tend to cancel, and the intensity tends to be zero. This extreme case occurs only if $A_1 = A_2$ and $\varphi_1 = \varphi_2$ (suppose $\varphi_1, \varphi_2 \in (0, 2\pi]$).



(a) *similar_to* (b) *spouse*

Figure 4: Distribution histogram of the phases of *_similar_to* in WN18RR and *spouse* in FB15k-237.

B Reasoning in Symmetric Relations

OpticE can infer symmetric, anti-symmetric, inversion and composition relations. We give the proof for symmetric relations here, and others are omitted in this paper.

For a symmetric relation r with two positive cases, (h, r, t) and (t, r, h), to satisfy the coherence condition of phase matching,

$$\begin{aligned} \varphi_h + \varphi_r - \varphi_t &= 2k_1\pi \\ \varphi_t + \varphi_r - \varphi_h &= 2k_2\pi, \end{aligned}$$

where k_1 and $k_2 \in \mathcal{Z}$. By adding the formulas above, we can obtain

$$\varphi_r = (k_1 + k_2)\pi = k\pi,$$

in which $k \in \mathcal{Z}$. Specifically, $\varphi_r = \pm\pi$ and 0 in OpticE.

The phase properties inherited from the rotation models are tested. Compared with translation models, e.g., TransE, rotation transformation models are able to represent symmetric relations. For instance, when (h, r, t) and (t, r, h) are both positive cases, it is trivial for TransE to set r to 0. However, for OpticE, the period of the phase takes effect with a phase $\varphi_r = k\pi$. This is verified in Figure 4, in which the phase of the symmetric relation *_similar_to* and *spouse* is distributed around $\pm\pi$ and 0.