

# Understanding Relation Temporality of Entities

Taesung Lee and Seung-won Hwang

Department of Computer Science and Engineering  
Pohang University of Science and Technology (POSTECH)  
Pohang, Republic of Korea  
{elca4u, swhwang}@postech.edu

## Abstract

This paper demonstrates the importance of relation equivalence for entity translation pair discovery. Existing approach of understanding relation equivalence has focused on using explicit features of co-occurring entities. In this paper, we explore latent features of temporality for understanding relation equivalence, and empirically show that the explicit and latent features complement each other. Our proposed hybrid approach of using both explicit and latent features improves relation translation by 0.16 F1-score, and in turn improves entity translation by 0.02.

## 1 Introduction

Understanding relations is important in entity tasks. In this paper, we illustrate such importance using named entity (NE) translation mining problem. Early research on NE translation used phonetic similarities, for example, to mine the translation ‘Mandelson’ → ‘曼德尔森’ [ManDeErSen] with similar sounds (Knight and Graehl, 1998; Wan and Verspoor, 1998). However, not all NE translations are based on transliterations, but they might be based on semantics (e.g., ‘WTO’ → ‘世贸组织’ [ShiMaoZuZhi]), or even arbitrary (e.g., ‘Jackie Chan’ → ‘成龙’ [ChengLong]).

To address this challenge, current state-of-the-art approaches build an entity graph for each language corpus, and align the two graphs by propagating the seed translation similarities (Figure 1) (Kim et al., 2011; You et al., 2012). For example, arbitrary translation pair such as (Jackie Chan, 成龙) can be obtained, if he is connected to his film ‘Drunken Master’ (醉拳) in both graphs. That is, we can propagate the seed translation similarity of (Drunken Master, 醉拳) to neighbor entities ‘Jackie Chan’ and ‘成龙’ in each graph.

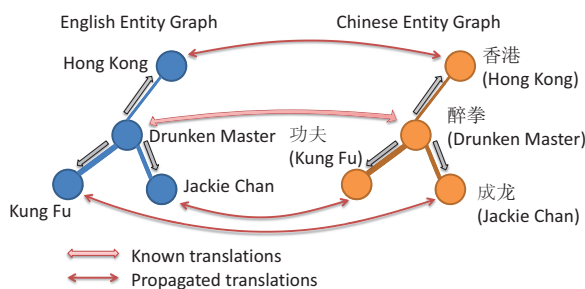


Figure 1: Entity translation by propagation.

When two graphs are obtained from parallel corpora, graphs are symmetric and “blind propagation” described above is effective. In contrast, Lee and Hwang (2013) propose “selective propagation” for asymmetric graphs, of comparing the semantics of relations. A key contribution of this paper is using relation temporality for determining relation equivalence. Existing work (Nakashole et al., 2012; Mohamed et al., 2011; Lee and Hwang, 2013) uses only co-occurring entity pairs, or explicit features (EF). For example, for a relation pay an official visit to, with a statement (Bush, pay an official visit to, China), an entity pair (Bush, China) is in the “support set”, which is a set of co-occurring entity pairs of pay an official visit to. When its support set is {(Bush, China), (Mandelson, Moscow), (Rice, Israel)}, and that of visit is {(Bush, China), (Rice, Israel), (Medvedev, Cuba)}, we can infer their semantic equivalence based on the set intersection: {(Bush, China), (Rice, Israel)}.

In contrast, we propose to explore corpus latent features (LF), to complement the sparsity problem of EF: Out of 158 randomly chosen correct relation translation pairs we labeled, 64% has only one co-occurring entity pair, which makes EF not very effective to identify these relation translations. Therefore, we leverage *relation temporality*, which is both orthogonal and complementary to existing efforts leveraging *entity temporality* (Kle-

mentiev and Roth, 2006; Kim et al., 2012; You et al., 2013). In particular, we discover three new challenges on using temporality for relation understanding in comparable corpora, which we discuss in detail in Section 3.2. Based on these challenges, we identify three new features for LF.

We observe the complementary nature of EF and LF, then propose a hybrid approach combining both features. Our new hybrid approach significantly improves the relation translation (0.16 higher F1-score than EF), and in turn improves the entity translation (0.02 higher F1-score).

## 2 Preliminary: Entity Translation by Selective Propagation

Selective propagation, leveraging the statements extracted from bilingual comparable corpora, can be summarized by several steps.

**STEP 1** Initialize entity translation function  $T_N^{(0)}$ .

**STEP 2** Build relation translation function  $T_R^{(t)}$  using  $T_N^{(t)}$ .

**STEP 3** Update entity translation function to acquire  $T_N^{(t+1)}$  using  $T_R^{(t)}$ .

**STEP 4** Repeat **STEP 2** and **STEP 3**.

For **STEP 1**, an existing method for entity translation is adopted. In our experiments, we use a non-selective (hence not requiring relation translations) propagation approach (You et al., 2012) with (Lam et al., 2007) for a base translation matrix. The focus of this paper is **STEP 2**, building the translation score  $T_R^{(t)}(r_E, r_C)$  of English relation  $r_E$  and Chinese relation  $r_C$ : We will discuss the detailed procedure of **STEP 2** and propose how to improve it in Section 3. **STEP 3** is the stage that selective propagation takes place.

**STEP 2** and **STEP 3** reinforce each other to improve the final entity translation function. While **STEP 3** is well-defined in (Lee and Hwang, 2013), to propagate entity translation scores when the relation semantics of the edges are equivalent, **STEP 2** has been restricted to the explicit feature, i.e., co-occurring entities or shared context. In clear contrast, by discovering novel latent features based on temporal properties, we can increase the accuracy of both entity and relation translations. Note that we omit  $t$  for readability in the following sections.

## 3 Relation Translation

In this section, we present our approaches to obtain relations of equivalent semantics across languages (e.g., *visit* → 访问). Formally, our goal is to build the relation translation score function  $T_R(r_E, r_C)$  for English relation  $r_E$  and Chinese relation  $r_C$ .

### 3.1 Baseline: Explicit Feature Approach (EF)

In this section, we briefly illustrate a baseline method EF (Lee and Hwang, 2013). As we mentioned in the introduction, traditional approaches leverage common co-occurring entity-pairs. This observation also holds in the bilingual environment by exploiting seed entity translations. For example, let us say that we have two extracted statements: (Bruce Willis, star in, The Sixth Sense) and (布鲁斯·威利斯 (Bruce Willis), 主演 (star in), 第六感 (The Sixth Sense)). Knowing a few seed entity translations using  $T_N$ , ‘Bruce Willis’ → ‘布鲁斯·威利斯’ and ‘The Sixth Sense’ → ‘第六感’, we can find *star in* and *主演* are semantically similar.

Specifically, we quantify this similarity based on the number of such common entity pairs that we denote as  $|H(r_E, r_C)|$  for an English relation  $r_E$  and a Chinese relation  $r_C$ . The existing approaches are variations of using  $|H(r_E, r_C)|$ . Our baseline implementation uses the one by (Lee and Hwang, 2013), and we refer the reader to the paper for formal definitions and processing steps we omitted due to the page limit.

Unfortunately, this approach suffers from sparsity of the common entity pairs due to the incomparability of the corpora and those entities that cannot be translated by  $T_N$ . Therefore, we leverage corpus latent features as an additional signal to overcome this problem.

### 3.2 Latent Feature Approach (LF)

#### Temporal Feature Discovery

We exploit the temporal distribution  $d[x](t)$  of textual element  $x$  during  $t$ -th week in statements; we count the occurrences of the element  $x$  on a weekly basis, and normalize them to observe  $\sum_t d[x](t) = 1$ . For example, Figure 2a shows the relation temporal distribution  $d[\text{visit}](t)$  against week  $t$ . Unlike entities, we can easily observe the dissimilarity of the temporal distributions of semantically equivalent relations. We identify the

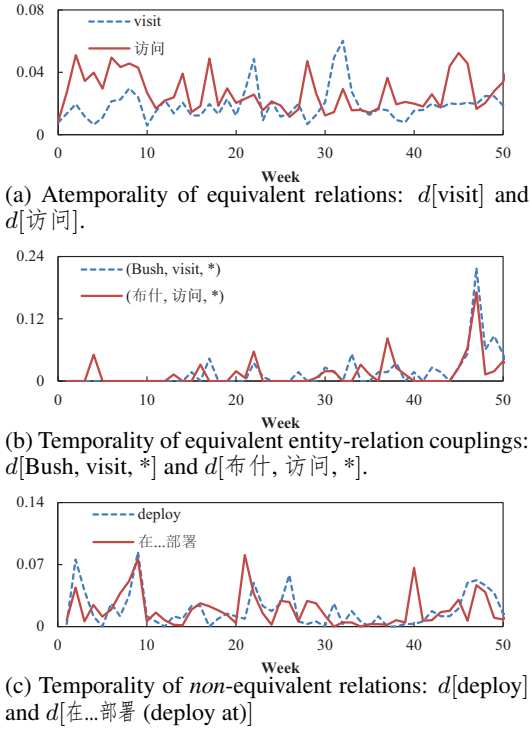


Figure 2: Temporal distributions of a relation, and a coupling.

three big challenges in exploiting the temporality in relation translations.

[C1] Considering temporal distributions  $d[r]$  of relations alone is not sufficient. For relations, such as visit, that involves diverse entities, the temporal distributions are highly noisy (Figure 2a).

To address the first challenge, we use a finer-granularity unit for observing the temporality. More specifically, we exploit a coupling of a relation and an entity:  $d[e, r, *]$  where  $e$  is an entity,  $r$  a relation, and  $*$  is a placeholder indicating that any noun phrase is accepted for the second argument of a statement.<sup>1</sup> As shown in Figure 2b,  $d[e, r, *]$  is more distinctive and hence a key clue to find semantically equivalent relations.

[C2] Considering entity-relation coupling distribution  $d[e, r, *]$  alone is not sufficient due to the domination of individual temporality. For example, Figure 3 shows *entity-dominating* entity-relation temporality. If an entity has a peak at some period (Figure 3a), most relations that are coupled with the entity would have a peak at the very same period (Figure 3b). This makes all relations that appear with this entity very similar to

<sup>1</sup>We use both  $d[e, r, *]$  and  $d[* , r, e]$  to measure the relation translation scores and leverage the average score. But in this section, we only use  $d[e, r, *]$  for readability.

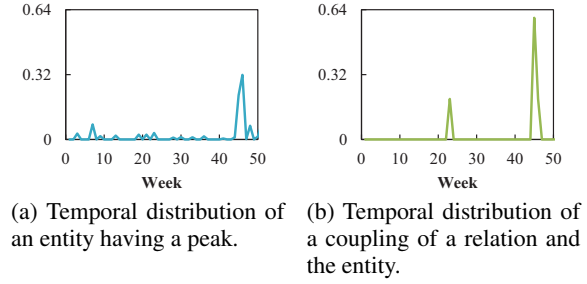


Figure 3: False positive peak of an entity-relation coupling.

each other regardless of semantics. To address this challenge, we use features to measure whether  $d[e, r, *]$  is too close to either of  $d[e]$  or  $d[r]$ .

[C3] Lastly, we have to eliminate false positives in relation temporality. To illustrate, two relations deploy and 在...部署 (deploy at) have similar temporal behaviors (Figure 2c). However, the first relation takes [person], but the second relation [location] for the second argument.

To address this, we check the common co-occurring entity pair of the relations. For example, we can obtain “Russia deployed an aircraft carrier”, but not “Russia deployed at (在...部署) an aircraft carrier”. Thus, we cannot acquire any common entity pair like (Russia, aircraft carrier) for deploy and 在...部署 (deploy at).

### Relation Similarity Computation

We compute the similarity of two relations  $r_E$  in English and  $r_C$  in Chinese using the following 2-steps.

- Compute the similarity  $S_{CP}(r_E, r_C, e_E, e_C)$  of temporal distributions of entity-relation couplings for each bilingual entity pair  $(e_E, e_C)$ .
- Compute the translation score  $T_R^{LF}(r_E, r_C)$  by aggregating the coupling similarities.

Considering the three challenges, we produce a list of features  $\{f_x(r_E, r_C, e_E, e_C)\}$  to measure the coupling similarity  $S_{CP}(r_E, r_C, e_E, e_C)$  as follows.

- [Base feature]  $f_{ET}$ :  $T_N(e_E, e_C)$ . The entity translation score obtained in the previous iteration or the seed entity translation score.
- [C1]  $f_{ER}$ :  $1 - JSD(d[e_E, r_E, *], d[e_C, r_C, *])$ . The temporal similarity of the couplings, where  $JSD(P, Q)$  is the Jensen-Shannon divergence of two distributions  $P$  and  $Q$ , defined as  $JSD(P, Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$ ,

with  $M = \frac{1}{2}(P + Q)$  and  $D(P||M) = \sum_i P(i) \log \frac{P(i)}{M(i)}$ .

- [C2]  $f_{D1,E}, f_{D2,E}, f_{D1,C}, f_{D2,C}$ :

$$JSD(d[e_E], d[e_E, r_E, *]), JSD(d[r_E], d[e_E, r_E, *])$$

$$JSD(d[e_C], d[e_C, r_C, *]), JSD(d[r_C], d[e_C, r_C, *])$$

Entity to entity-relation distribution difference (D1) and relation to entity-relation distribution difference (D2), for English and Chinese respectively.

- [C3]  $f_{EX}$ : The existence of a common entity pair using the seed entity translations (boolean). That is,  $f_{EX} = 1$  if  $|H(r_E, r_C)| \geq 1$ , and  $f_{EX} = 0$  otherwise.

Additionally, we use the following features to consider absolute frequencies  $freq(\cdot)$  of textual elements as well because 1) we are more confident with more evidence and 2) in the comparable corpora, the equivalent elements are likely to show similar frequencies.

- $f_{FW,E}, f_{FW,C}$ :  $\mathcal{S}(freq(e_E, r_E))$  and  $\mathcal{S}(freq(e_C, r_C))$ .  $\mathcal{S}(x)$  is a normalization function, for which we use a sigmoid function over a linear transformation of  $x$ .
- $f_{FS1}$  and  $f_{FS2}$ :

$$\frac{\min(freq(e_E, r_E), freq(e_C, r_C))}{\max(freq(e_E, r_E), freq(e_C, r_C))},$$

$$\frac{\min(freq(r_E), freq(r_C))}{\max(freq(r_E), freq(r_C))}$$

With these features, we measure the similarity of a pair of couplings as follows.

$$S_{CP}(r_E, r_C, e_E, e_C) = \prod_x f_x(r_E, r_C, e_E, e_C) \quad (1)$$

By aggregating coupling similarities, we measure the translation score of two relations:

$$T_R^{LF}(r_E, r_C) = \sum_{(e_E, e_C) \in T} S_{CP}(r_E, r_C, e_E, e_C) \quad (2)$$

where  $T = \{(e_E, e_C) | T_N(e_E, e_C) \geq \theta\}$  with  $\theta = 0.6$ , a set of translation pairs obtained in the seeds or previous iteration such as (Bush, 布什).

We normalize the obtained function values for each English relations using the top- $k$  Chinese translations. That is, for  $(r_E, r_C)$ , we redefine the score as  $T_R^{LF}(r_E, r_C) / \sum_{i \in [1, k]} T_R^{LF}(r_E, r_C^{rank_i})$  where  $r_C^{rank_i}$  is the  $i$ -th rank Chinese relation for  $r_E$  by Equation 2. We empirically set  $k = 4$ .

English	LF	EF
visit	访问 (visit)	访问 (visit)
support	向...提供 (provide to ...)	-
ratify	讨论 (discuss) <sup>2</sup>	批准 (ratify)

Table 1: Examples of relation translations.

Method	Person			Organization		
	P.	R.	F1	P.	R.	F1
LF+EF	<b>0.84</b>	<b>0.80</b>	<b>0.82</b>	<b>0.60</b>	<b>0.52</b>	<b>0.56</b>
EF	0.81	0.79	0.80	0.56	<b>0.52</b>	0.54
Seed	0.80	0.77	0.78	0.49	0.44	0.46
PH+SM	0.59	0.59	0.59	0.29	0.29	0.29

Table 2: Entity translation comparison.

### 3.3 Hybrid Approach LF+EF

We find that LF and EF are complementary. Table 1 shows the examples of relations and their translations. In general, LF can translate more relations (e.g., support and capture). However, in some cases like ratify, highly *related* relations may induce noise. That is, we always 讨论 (discuss) before we 批准 (ratify) something and hence the temporal behavior of 讨论 (discuss) is also very similar to that of ratify. On the other hand, it can be correctly translated using EF.

Thus, we produce the hybrid relation translation, and we empirically set  $\lambda = 0.4$ :

$$T_R^{LF+EF}(r_E, r_C) = \lambda T_R^{LF}(r_E, r_C) + (1 - \lambda) T_R^{EF}(r_E, r_C) \quad (3)$$

## 4 Evaluation

In this section, we evaluate the proposed approach on the entity translation task and the relation translation task. We extract English and Chinese statements from news articles in 2008 by Xinhua news who publishes news in both English and Chinese, which were also used by Lee and Hwang (2013). The number of English articles is 100,746, and that of Chinese articles is 88,031. As we can see from the difference in the numbers of the documents, the news corpora are not direct translations, but they have asymmetry of entities and relations.

### 4.1 Entity Translation

In this section, we present experimental settings and results on translating entities using our proposed approaches. To measure the effectiveness,

<sup>2</sup>The correct translation 批准 (ratify) is ranked second.

Methods	Precision	Recall	F1
LF+EF	0.37	<b>0.44</b>	<b>0.40</b>
LF	0.26	0.25	0.26
EF	<b>0.41</b>	0.17	0.24

Table 3: Relation translation comparison.

we use a set of gold standard entity translation pairs which consist of 221 person entities and 52 organization entities. We measure precision, recall, and F1-score based on the returned translation pairs for each English entity as it is done in (Lee and Hwang, 2013).

We compare our hybrid approach, which is denoted by LF+EF with EF (Lee and Hwang, 2013), a combined approach PH+SM of phonetic similarity and letter-wise semantic translation for better accuracy for organizations (Lam et al., 2007), and the seed translations Seed that we adopt (You et al., 2012) with PH+SM as a base translation matrix.<sup>3</sup> We process one iteration of the entire framework (STEP 1-3) for both LF+EF and EF.

Table 2 shows the comparison of the methods. Our proposed approach LF+EF shows higher F1-score than the baselines. In particular, our approach outperforms EF. For example, ‘Matthew Emmons’ is a lesser known entity, and we have only few statements mentioning him in the corpora. The corpus explicit feature EF alone cannot translate the relation win and, in turn, ‘Matthew Emmons’. However, LF+EF translates him correctly into 马修·埃蒙斯 through the relation win.

## 4.2 Relation Translation

This section considers the relation translation task. Each relation translation method translates an English relation  $r_E$  into a list of Chinese relations, and we check whether a Chinese relation  $r_C$  with the highest translation score is the correct translation. We consider the relation translation is correct when the semantics are equivalent. For example, 去 (leave for/go to) is a correct translation of leave for, but 离开 (leave) is *not*. Total 3342 English-Chinese relation translation pairs returned by our method and the baselines are randomly shown and labeled. Out of 3342 pairs, 399 are labeled as correct.

<sup>3</sup>Our results leveraging relational temporality outperforms the reported results using entity temporality on the same data set. The two approaches using temporality are orthogonal and can be aggregated, which we leave as our future directions.

Eng. Rel.	C1	C1+C2	C1+C2+C3	EF
visit	15	4	1	1
drop	21	14	1	-
capture	6	4	1	-

Table 4: Rank of correct relation translation. The symbol ‘-’ indicates no correct translation.

Table 3 shows the comparisons of LF, EF and their hybrid LF+EF. We can clearly see that LF shows higher recall than EF while EF shows higher precision. As we emphasized in Section 3.3, we can see their complementary property. Their hybrid LF+EF has both high precision and recall, thus has the highest F1-score.

Note that the absolute numbers (due to the harsh evaluation criteria) may look low. But the top translations are still relevant (e.g., fight is translated to 驻 (deploy troops)). In addition, the lower ranked but correct relation translations also affect entity translation. Therefore, even lower-performing EF boosted the entity translations, and in effect, our approach could achieve higher F1-score in the entity translation task.

To illustrate the detailed effects of the corpus latent features, Table 4 shows the ranks of correct Chinese translations for English relations by methods using selected features for the challenges. For comparison, the ranks of the correct translations when using EF are shown. Our approach using the entity-relation coupling similarity feature for [C1] alone often cannot find the correct translations. But using all features removes such noise.

## 5 Conclusion

This paper studied temporality features for relation equivalence. With the proposed features, we devised a hybrid approach combining corpus latent and explicit features with complementary strength. We empirically showed the effectiveness of our hybrid approach on relation translation, and it, in turn, improved entity translation.

## Acknowledgments

This research was supported by the MSIP (The Ministry of Science, ICT and Future Planning), Korea and Microsoft Research, under IT/SW Creative research program supervised by the NIPA(National IT Industry Promotion Agency). (NIPA-2013-H0503-13-1009).

## References

- Jinhan Kim, Long Jiang, Seung-won Hwang, Young-In Song, and Ming Zhou. 2011. Mining entity translations from comparable corpora: a holistic graph mapping approach. In *Proc. 20<sup>th</sup> ACM International Conference on Information and Knowledge Management (CIKM 2011)*, pages 1295–1304. ACM.
- Jinhan Kim, Seung-won Hwang, Long Jiang, Young-In Song, and Ming Zhou. 2012. Entity translation mining from comparable corpora: Combining graph mapping with corpus latent features. In *IEEE Transactions on Knowledge and Data Engineering*, pages 1787–1800.
- Alexandre Klementiev and Dan Roth. 2006. Named entity transliteration and discovery from multilingual comparable corpora. In *Proc. 8<sup>th</sup> Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006)*, pages 82–88. Association for Computational Linguistics.
- Kevin Knight and Jonathan Graehl. 1998. Machine transliteration. *Computational Linguistics*, 24(4):599–612, December.
- Wai Lam, Shing-Kit Chan, and Ruizhang Huang. 2007. Named entity translation matching and learning: With application for mining unseen translations. *ACM Transactions on Information Systems*, 25(1), February.
- Taesung Lee and Seung-won Hwang. 2013. Bootstrapping entity translation on weakly comparable corpora. In *Proc. 51<sup>st</sup> Annual Meeting on Association for Computational Linguistics (ACL 2013)*. Association for Computational Linguistics.
- Thahir Mohamed, Estevam Hruschka, and Tom Mitchell. 2011. Discovering relations between noun categories. In *Proc. 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 1447–1455. Association for Computational Linguistics.
- Ndapandula Nakashole, Gerhard Weikum, and Fabian M. Suchanek. 2012. PATTY: A Taxonomy of Relational Patterns with Semantic Types. In *Proc. 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*. Association for Computational Linguistics.
- Stephen Wan and Cornelia Maria Verspoor. 1998. Automatic English-Chinese name transliteration for development of multilingual resources. In *Proc. 36<sup>th</sup> Annual Meeting on Association for Computational Linguistics (ACL 1998) and 17<sup>th</sup> International Conference on Computational Linguistics (COLING 1998)*, pages 1352–1356. Association for Computational Linguistics.
- Gae-won You, Seung-won Hwang, Young-In Song, Long Jiang, and Zaiqing Nie. 2012. Efficient entity translation mining: A parallelized graph alignment approach. *ACM Transactions on Information Systems*, 30(4):25:1–25:23, November.
- Gae-won You, Young-rok Cha, Jinhan Kim, and Seung-won Hwang. 2013. Enriching entity translation discovery using selective temporality. In *Proc. 51<sup>st</sup> Annual Meeting on Association for Computational Linguistics (ACL 2013)*, pages 201–205. Association for Computational Linguistics.