UM4: Unified Multilingual Multiple Teacher-Student Model for Zero-Resource Neural Machine Translation

Jian Yang[1*], Yuwei Yin[2*], Shuming Ma[2], Dongdong Zhang[2], Shuangzhi Wu[3], Hongcheng Guo[2], Zhoujun Li[1], Furu Wei[2]

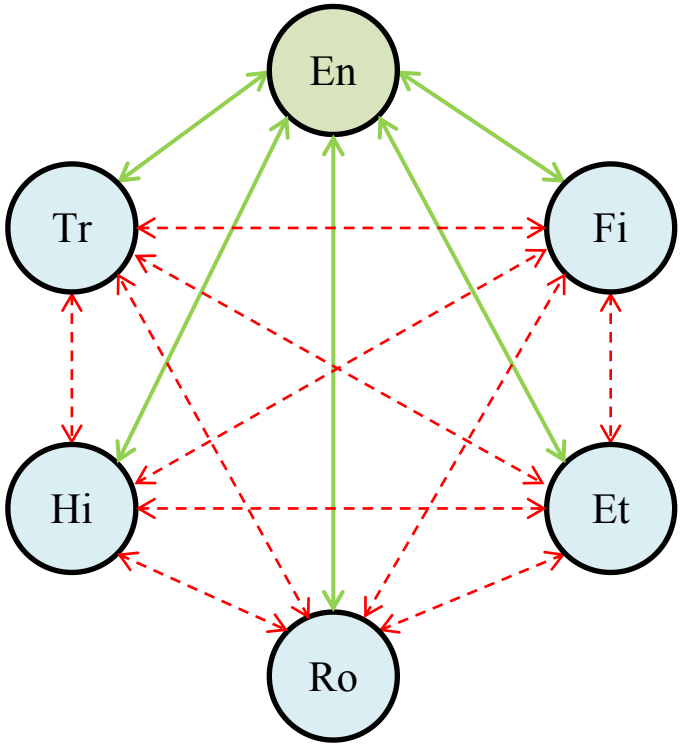
[1]State Key Lab of Software Development Environment,

Beihang University, Beijing, China

[2]Microsoft Research Asia

[3]Tencent Cloud Xiaowei

Background: zero-resource translation



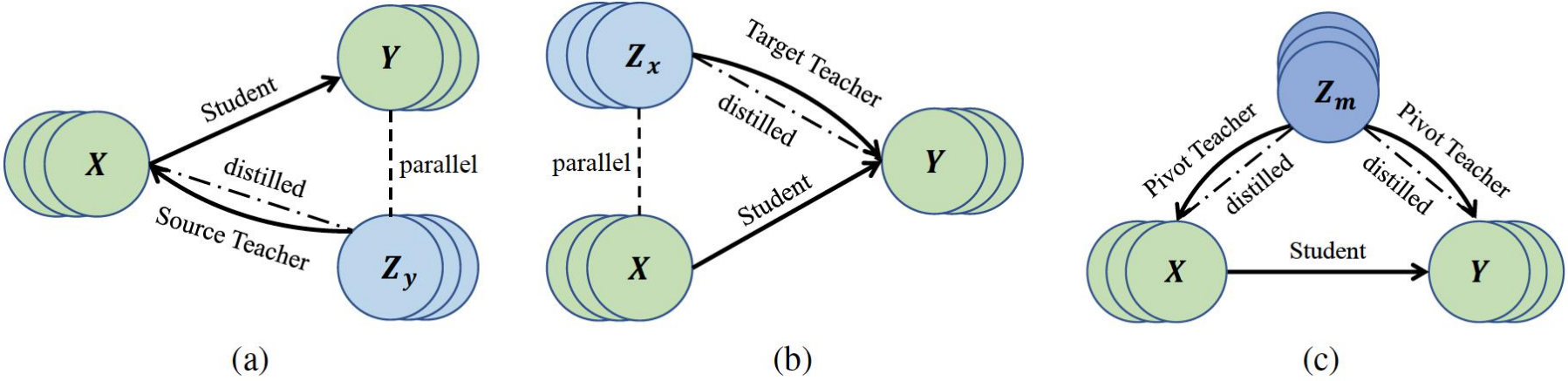
English(En)-centric parallel corpora

6 languages, $6 \times (6 - 1) = 30$ directions

√ 5 En↔X parallel corpora

× 25 parallel corpora of all other directions

Method: UM4

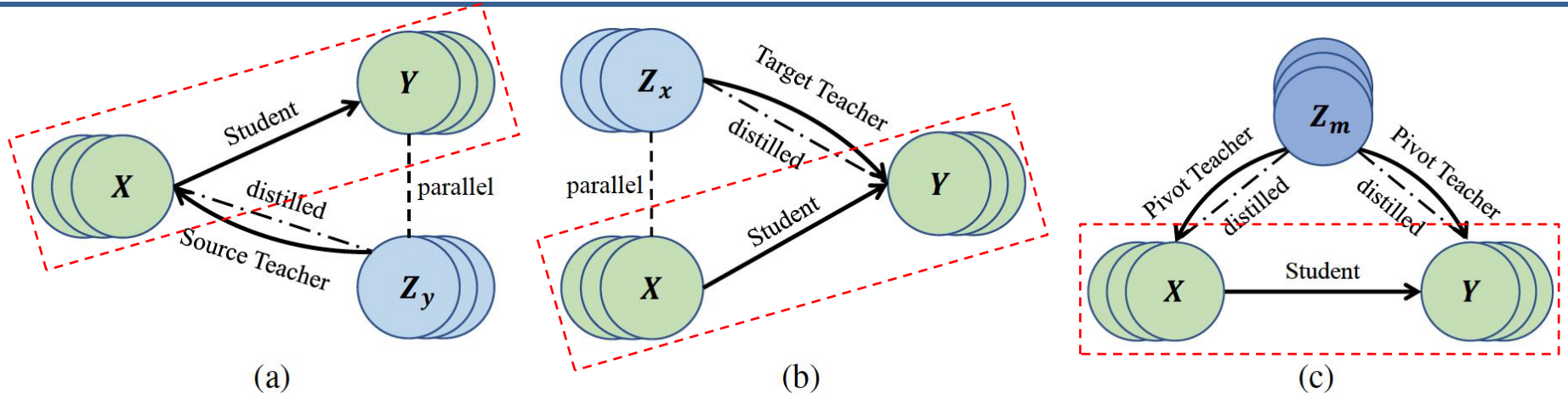


Source Teacher

Target Teacher

Pivot Teacher

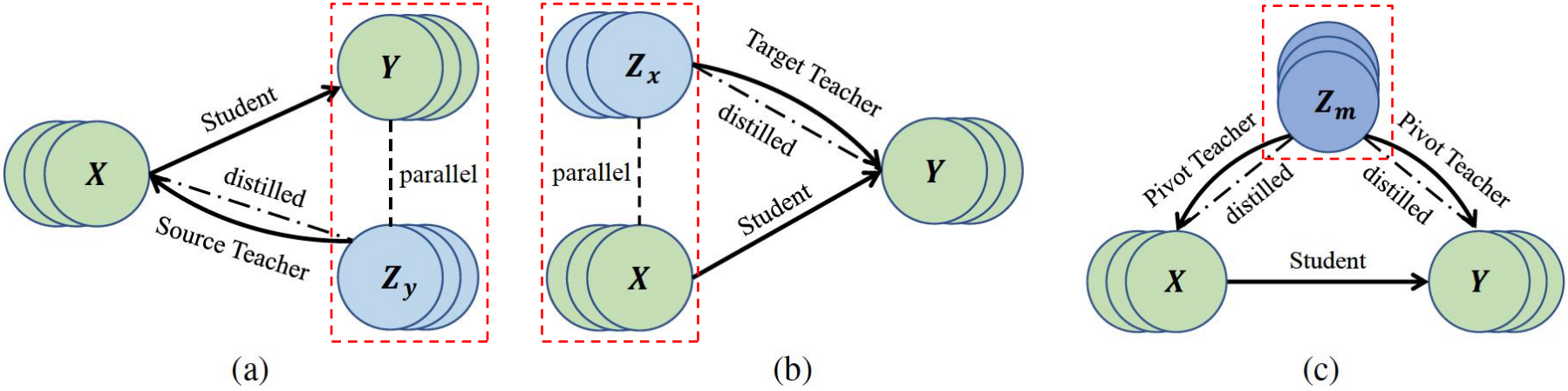
Method: UM4



Source Teacher | Target Teacher | Pivot Teacher

Goal: tranlate X→Y. (problem: lack of X-Y parallel data)

Method: UM4

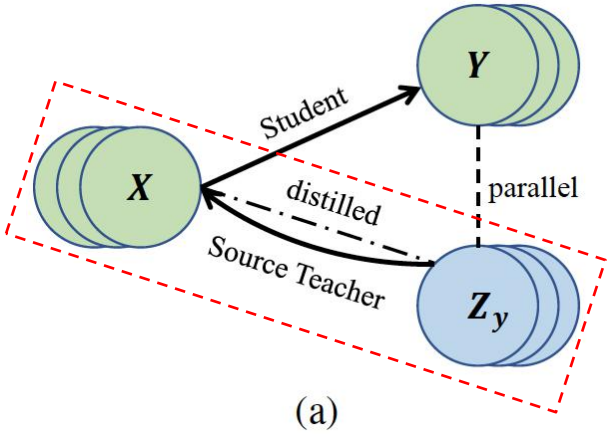


Source Teacher Target Teacher Pivot Teacher

Goal: tranlate X→Y. (problem: lack of X-Y parallel data)

Available parallel data: Y-Z_y & X-Z_x ; Available monolingual data: Z_m

Method: UM4 - Source Teacher



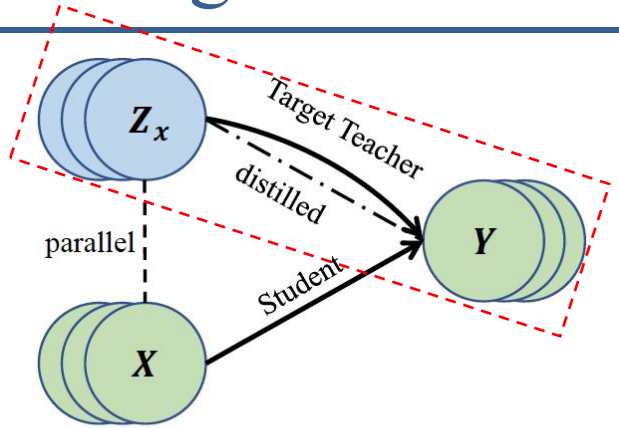
(a)

Source Teacher

Goal: tranlate X→Y. (problem: lack of X-Y parallel data)
Available parallel data: Y-Z_y & X-Z_x ; Available monolingual data: Z_m

Source Teacher: translate Z_y→pseudo X. (pseudo X, real Y) pair

Method: UM4 - Target Teacher



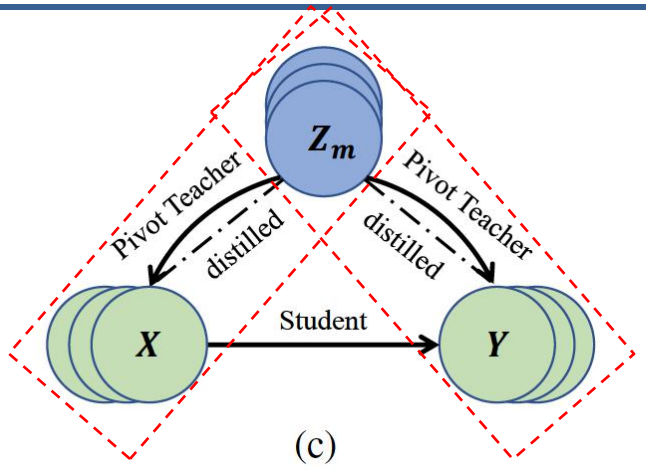
(b)

Target Teacher

Goal: tranlate X→Y. (problem: lack of X-Y parallel data)
Available parallel data: Y-Z_y & X-Z_x ; Available monolingual data: Z_m

Target Teacher: translate Z_x→pseudo Y. (real X, pseudo Y) pair

Method: UM4 - Pivot Teacher



Pivot Teacher

Goal: tranlate X→Y. (problem: lack of X-Y parallel data)
Available parallel data: Y-Z_y & X-Z_x ; Available monolingual data: Z_m

Pivot Teacher: translate Z_m→pseudo X/Y. (pseudo X, pseudo Y) pair

Experiment: WMT benchmark

$9 \times (9 - 1) = 72$ all directions, including 16 original parallel pairs En→X & X→En

$8 \times (8 - 1) = 56$ **zero-resource** translation directions

	En	Fr	Cs	De	Fi	Et	Ro	Hi	Tr
En	-	10.2K	7.8K	9.6K	2.7K	2.0K	9.4K	2.2K	9.7K
Fr	10.2K	-	7.7K	9.0K	2.5K	2.0K	8.8K	2.1K	8.8K
Cs	7.8K	7.7K	-	7.0K	2.0K	1.2K	7.0K	1.7K	7.0K
De	9.6K	9.0K	7.0K	-	2.6K	1.9K	8.6K	2.0K	8.7K
Fi	2.7K	2.5K	2.0K	2.6K	-	0.7K	2.4K	0.7K	2.3K
Et	2.0K	2.0K	1.2K	1.9K	0.7K	-	1.9K	0.4K	1.7K
Ro	9.4K	8.8K	7.0K	8.6K	2.4K	1.9K	-	2.0K	8.4K
Hi	2.2K	2.1K	1.7K	2.0K	0.7K	0.4K	2.0K	-	1.9K
Tr	9.7K	8.8K	7.0K	8.7K	2.3K	1.7K	8.4K	1.9K	-

Conclusion

- ✓ In this work, we propose **a novel method** called Unified Multilingual Multiple teacher-student Model for NMT (**UM4**) to ameliorate the translation of **zero-resource directions**.
- ✓ Our method **unifies** the **source-teacher** model, **target-teacher** model, and **pivot-teacher** model to guide the multilingual source→target student model, alleviating the **error propagation** problem caused by **two-pass translation**.
- ✓ Experimental results on the **multilingual dataset** of the WMT benchmark corroborate the effectiveness of our method in **leveraging the distilled knowledge from the unified teachers**.

- ✓ Our **code** and **data** have been released
  - ➢ https://github.com/YuweiYin/UM4

Thanks!