Identifying Semantic Divergences in Parallel Text without Annotations

Yogarshi Vyas and **Xing Niu** and **Marine Carpuat**
Department of Computer Science
University of Maryland

*Accepted into NAACL'18

Motivation

- Intuition: Sentence alignment does not necessarily imply semantic equivalence (in current translation corpora).
- Purposes of this paper: provide empirical evidence that (1) semantic divergences exist in parallel corpora; (2) the divergences matter for downstream applications, e.g., MT.
- Contribution: a framework which can (1) detect divergence in parallel sentences, (2) without manual annotation.

Cross-Lingual Semantic Similarity Model

- Very Deep Pairwise Interaction (VDPWI) model*
  - ➢ Used to detect semantic textual similarity (STS) between English sentence pairs.
- Adapt the model to the cross-lingual task.
- Five components:
  - ➢ Bilingual word embeddings;
  - ➢ BiLSTM for contextualizing words;
  - ➢ Word similarity cube: pairwise word scores;
  - ➢ Similarity focus layer: reweight word pairs;
  - ➢ Deep convolutional network.

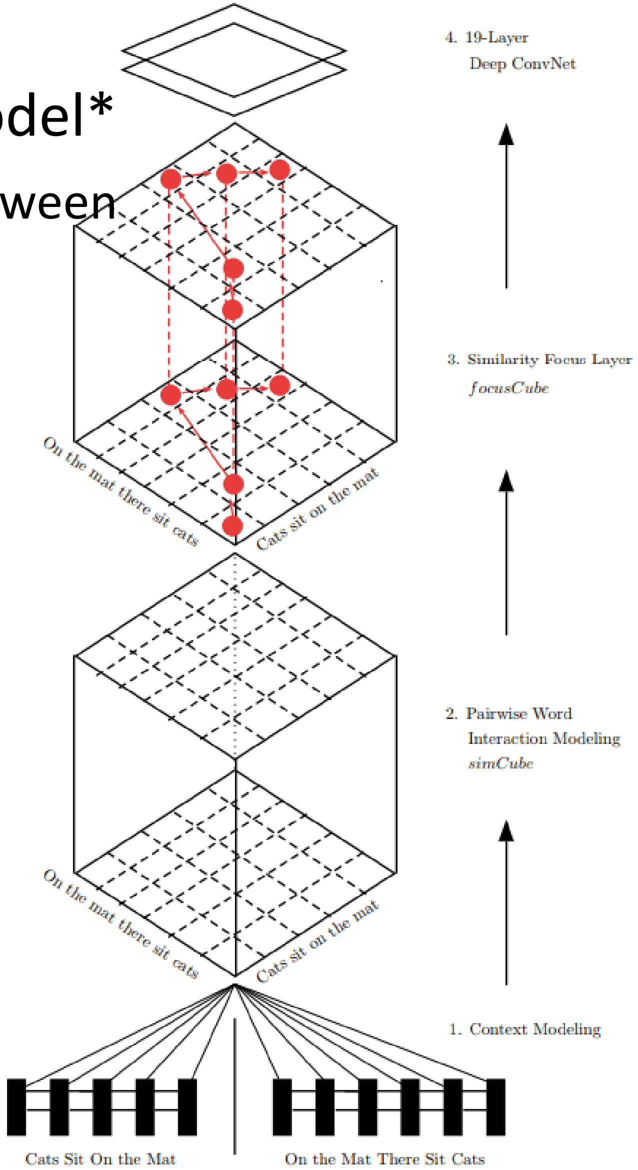


Figure 1: Our end-to-end neural network model, consisting of four major components.

*: Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement. –NAACL’16

Noisy Synthetic Supervision

- The model is trained to minimize KL divergence between the output similarity score and gold similarity score.
- Use parallel sentences as positive examples. $\{(e_i, f_i)\ \forall i\}$
- Candidate negative examples: take the Cartesian product of the two sides of the positive examples. $\{(e_i, f_j) \forall i, j \text{ s.t. } i \neq j\}$
- Filtering: only retain pairs that have close length (at most 1:2), and have enough words (at least half) which have a translation in the other sentence.

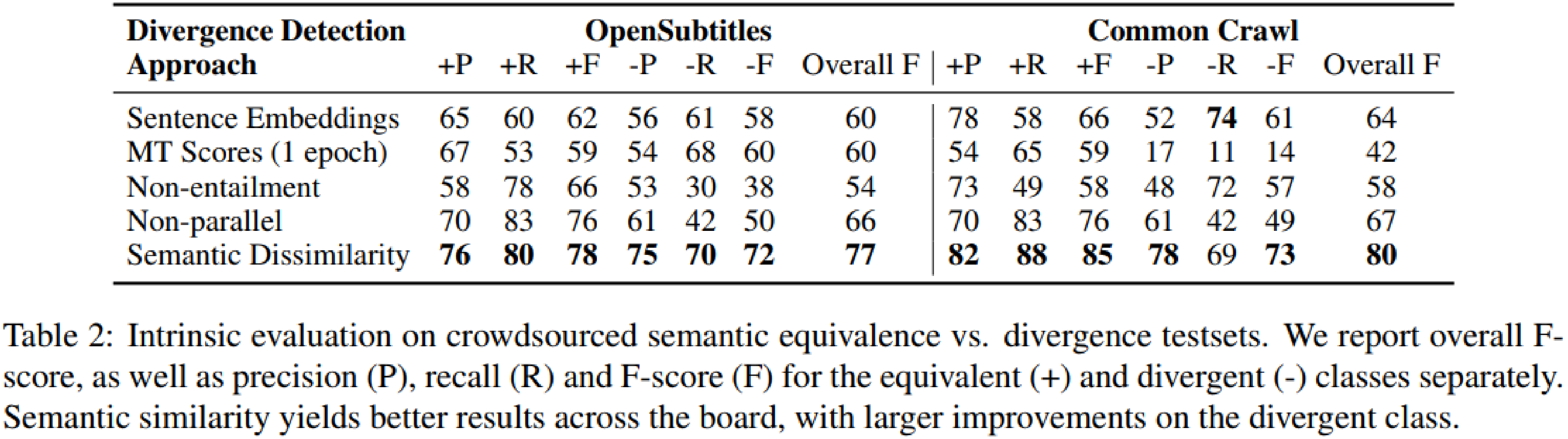
Crowdsourcing Divergence Judgments

- Annotations of English-French sentence pairs to construct test beds for evaluation.
- Datasets: *OpenSubtitles corpus* and *Common Crawl corpus*
- "the French and English text convey the same information."
- Analysis: 43.6% divergent examples in OpenSubtitles, and 38.4% in Common Crawl.

Divergence Detection Evaluation

- Baseline Models: (1)Parallel vs. Non-parallel Classifier, (2) Neural MT, (3) Bilingual Sentence Embeddings, (4) Textual Entailment Classifier;

- Classification on the two parallel corpora:

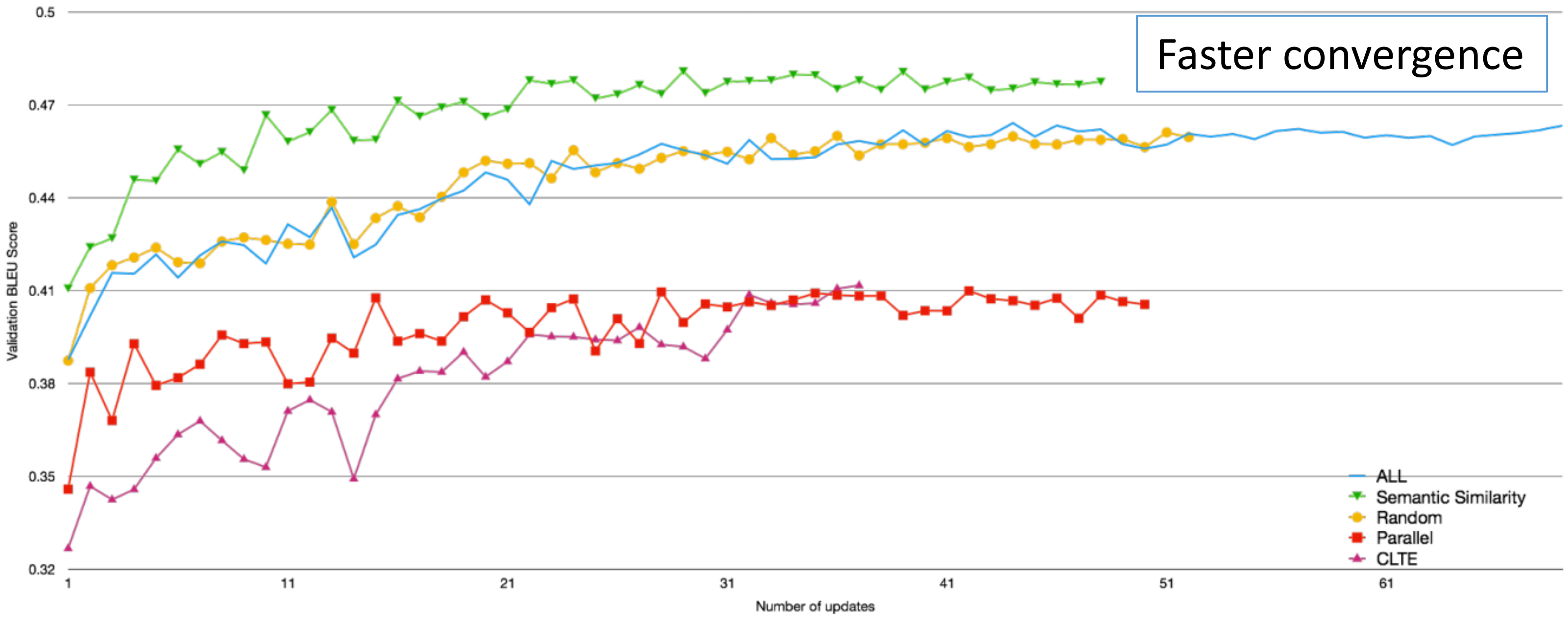


Divergence Detection			**OpenSubtitles**								**Common Crawl**				
Approach	+P	+R	+F	-P	-R	-F	Overall F	+P	+R	+F	-P	-R	-F	Overall F
Sentence Embeddings	65	60	62	56	61	58	60	78	58	66	52	**74**	61	64
MT Scores (1 epoch)	67	53	59	54	68	60	60	54	65	59	17	11	14	42
Non-entailment	58	78	66	53	30	38	54	73	49	58	48	72	57	58
Non-parallel	70	83	76	61	42	50	66	70	83	76	61	42	49	67
Semantic Dissimilarity	**76**	**80**	**78**	**75**	**70**	**72**	**77**	**82**	**88**	**85**	**78**	69	**73**	**80**

Table 2: Intrinsic evaluation on crowdsourced semantic equivalence vs. divergence testsets. We report overall F-score, as well as precision (P), recall (R) and F-score (F) for the equivalent (+) and divergent (-) classes separately. Semantic similarity yields better results across the board, with larger improvements on the divergent class.

Machine Translation Evaluation

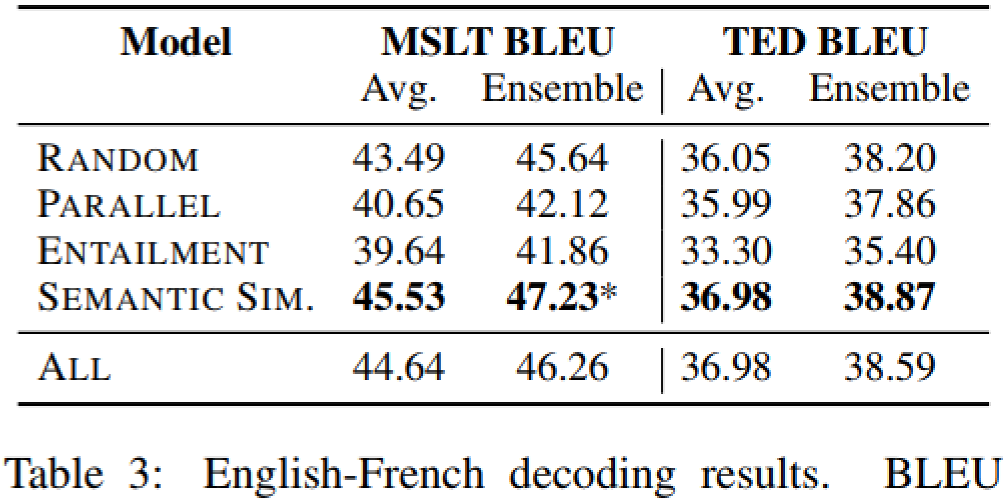
- Data Selection: select the least divergent examples. (50% in English-French and 90% in Vietnamese-English)
- NMT: RNNSearch



Faster convergence

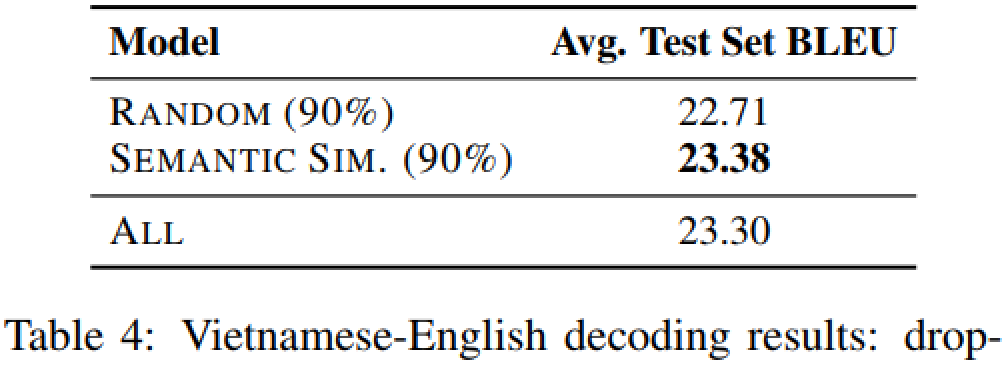
Figure 1: Learning curves on the validation set for English-French models (mean of 3 runs/model). The SEMANTIC SIMILARITY model outperforms other models throughout training, including the one trained on all data.

Translation Results



Model	MSLT BLEU		TED BLEU	
	Avg.	Ensemble	Avg.	Ensemble
RANDOM	43.49	45.64	36.05	38.20
PARALLEL	40.65	42.12	35.99	37.86
ENTAILMENT	39.64	41.86	33.30	35.40
SEMANTIC SIM.	**45.53**	**47.23***	**36.98**	**38.87**
ALL	44.64	46.26	36.98	38.59

Table 3: English-French decoding results. BLEU



Model	Avg. Test Set BLEU
RANDOM (90%)	22.71
SEMANTIC SIM. (90%)	**23.38**
ALL	23.30

Table 4: Vietnamese-English decoding results: drop-