UAlberta at SemEval-2024 Task 1: A Potpourri of Methods for Quantifying Multilingual Semantic Textual Relatedness and Similarity

Ning Shi, Senyu Li, Guoqing Luo, Amirreza Mirzaei, Ali Rafiei, Jai Riley, Hadi Sheikhi, Mahvash Siavashpour, Mohammad Tavakoli, Bradley Hauer, Grzegorz Kondrak

Department of Computing Science
University of Alberta, Edmonton, Canada

Semantic Textual Relatedness (STR) and Similarity (STS)

STR is a broad term for measuring the **degree of commonality** between pairs of sentences.

STS measures the degree in which pairs of sentences are **close in meaning.**

Semantic Textual Relatedness (STR) and Similarity (STS)

When I tried again, I was able to juggle.
When I went back to it, I was able to juggle!

High relatedness
High similarity

Old car driving down the road.
Two old women enjoying at a gathering.

Low relatedness
Low similarity

*SemEval 2024 Task 1 English Train

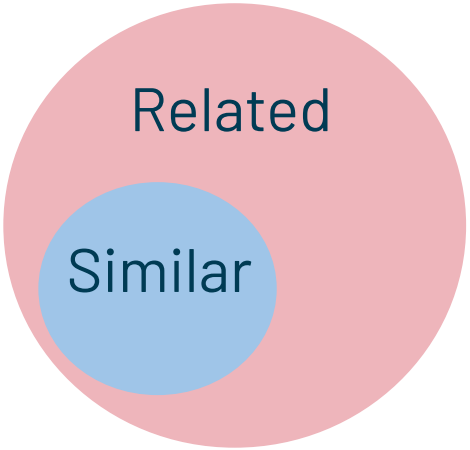
Hypothesis 1

Similarity is a special case of relatedness.

For example*:

> And in the United States, **we're considered** Mexican.
>
> And in the United States, **we are considering** Mexicans.

High relatedness but low similarity.



*SemEval 2017 Task 1 Track 4b

Hypothesis 2

Relatedness and similarity are **preserved under translation.**

It is better **known as a walk** . **It is** also **known as a walk** .	0.88	
It is better **known as a walk** . También **se le conoce como paseo** .	0.88	Spanish
It is better **known as a walk** . **Dit staan** ook **bekend as 'n stap** .	0.88	Afrikaans

Methods

Explicit Semantic	Extrinsic	Distributional	Large Language Models
Create and compare semantic representations of each inputted sentence	Use the output of systems designed for other semantic tasks	Create and compare embeddings from PLMs	Prompting or combining multiple model outputs

Methods

Explicit Semantic	Extrinsic	Distributional	Large Language Models
Word Overlap **(WO)** Concept Overlap **(CO)** Abstract Meaning Representation **(AMR)**	Paraphrase Identification **(PI)** Natural Language Inference **(NLI)**	**Embed-B** **Embed-R**	**Prompt** **Fusion** **Fine-tune**

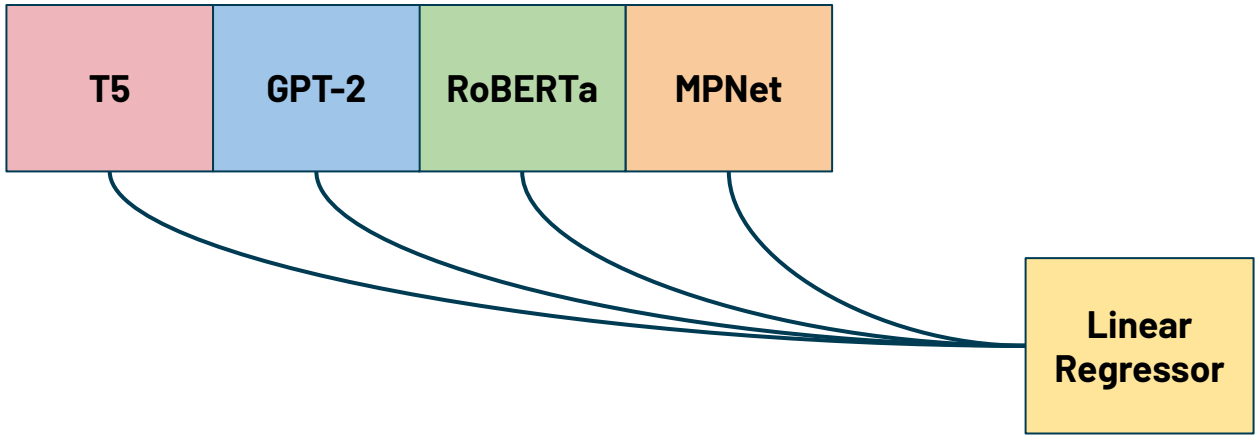
Methods

Explicit Semantic	Extrinsic	Distributional	Large Language Models
WO: Python Libraries **CO:** AMuSE-WSD **AMR:** Sapienza API	**PI:** RoBERTa & fine-tuned classifier **NLI:** RoBERTa with NLI Classifier	**Embed-B:** BERT **Embed-R:** RoBERTA	**Prompt:** ChatGPT **Fusion:** Open-source LLMs **Fine-tune:** T5, GPT2, RoBERTa, MPNet

Ensemble

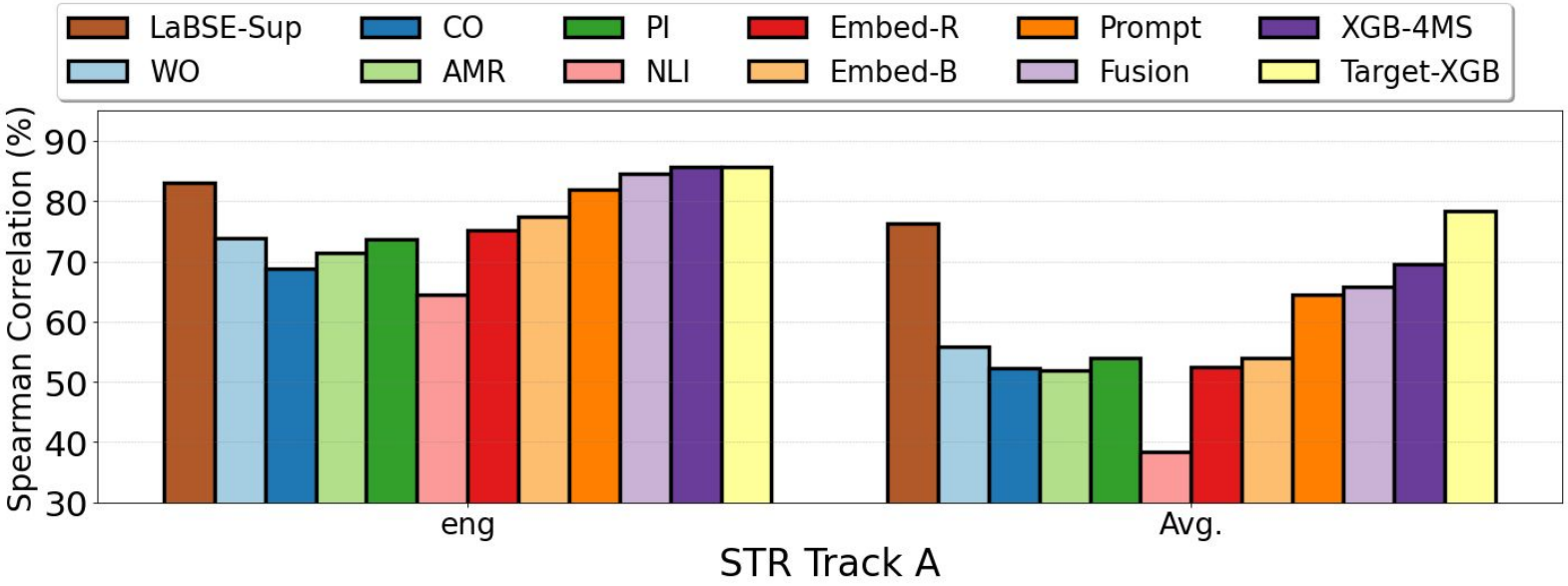
Our **best results** are reported from a **regression ensemble system** involving the **4 fine-tuned models.**

- Treat **each score** as a **feature in a linear regressor.**



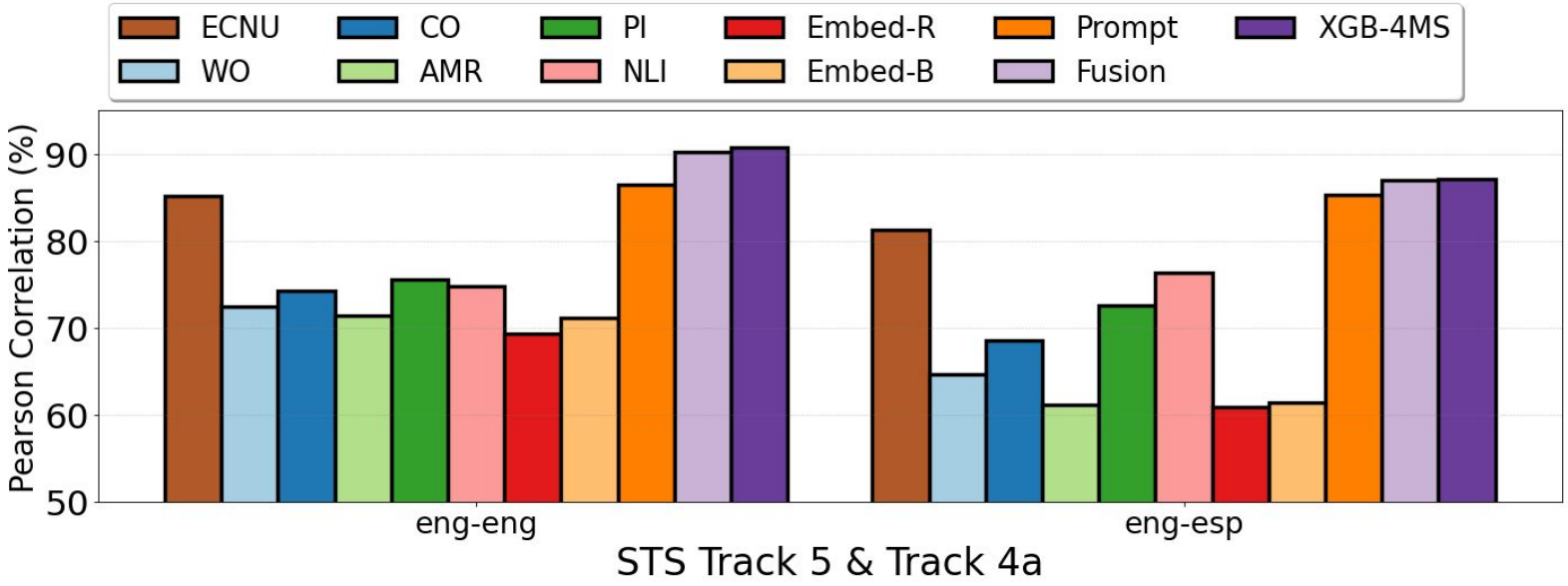
STR Results

Achieved SOTA results for English.



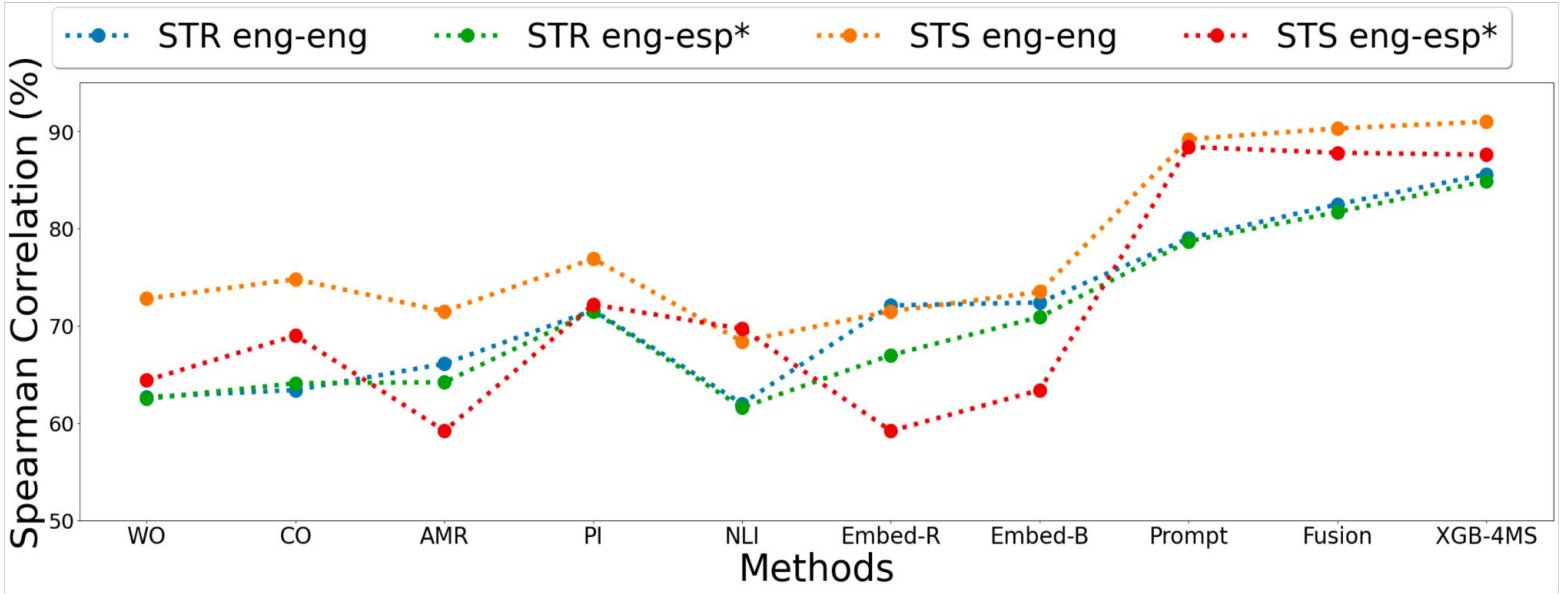
STS Results

STS dataset from SemEval 2017 Task 1 with ECNU being the best recorded method.

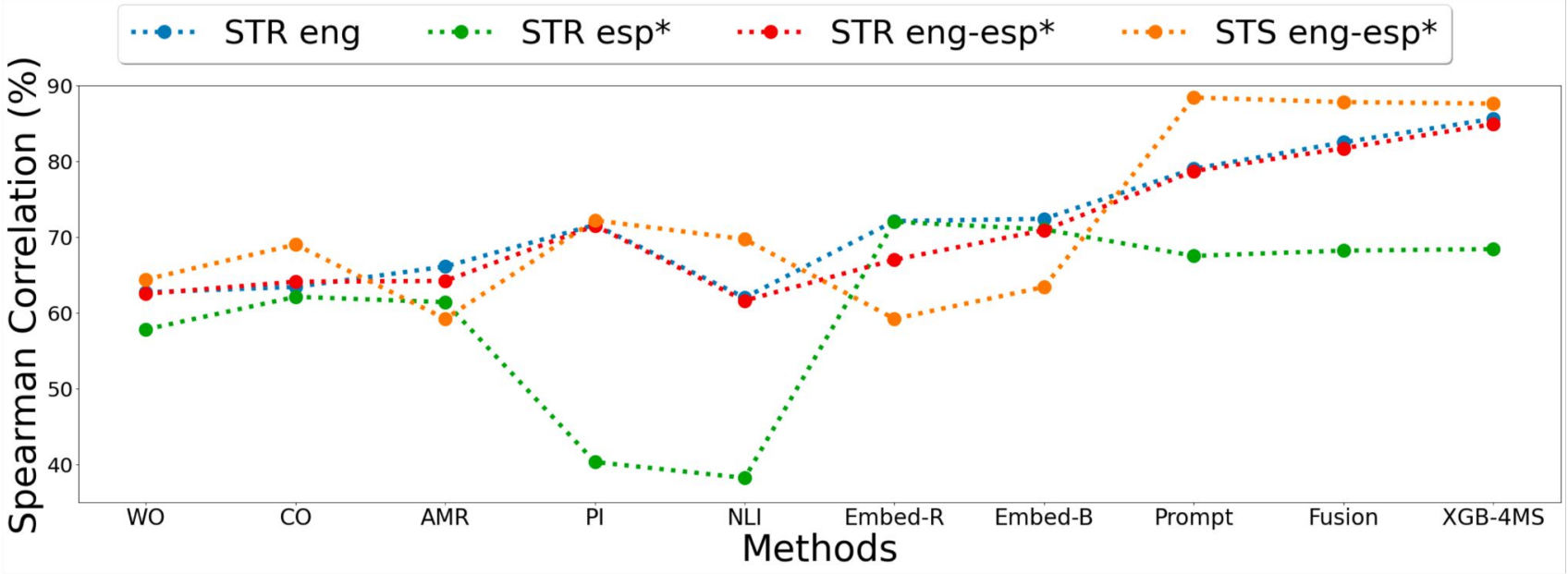


STR vs STS Results

High correlation between performance of methods on STR and STS datasets.



Mono-Lingual vs Cross-Lingual



Conclusions

- **Ensembled a variety of methods** on two sentence-level semantic tasks in mono-lingual and cross-lingual conditions.
- Achieved **SOTA results for English** and **top 3** performance for **16** of the language/track settings.
- Provided evidence for **two hypotheses:**
  1. Semantic **similarity is a special case of semantic relatedness.**
  2. Both similarity and relatedness are **preserved under translation.**

github.com/UAlberta-NLP/SemEval2024-STR