FACULTY OF ARTS
AND PHILOSOPHY

Contact

Shijie.Liu@UGent.be
https://lt3.ugent.be/people/shijie-liu/
@Ethan_LSJ_CN

LANGUAGE AND TRANSLATION TECHNOLOGY TEAM (LT3)

Shijie Liu, Els Lefever, and Véronique Hoste

TERMINOLOGY ANNOTATION IN THE MARITIME SAFETY DOMAIN: EVALUATING BERT FINE-TUNING AND GENERATIVE LANGUAGE MODEL PREDICTIONS

Background

- Automatic terminology extraction (ATE) has not been investigated in maritime safety domain, despite its specialized terminology characteristics

Previous studies

- Finetune pre-trained model has shown effectiveness in ATE research through transfer learning (Tran, 2024)
- LLMs demonstrate promising performance in specialized NLP tasks, including zero-shot learning scenarios

Research questions

- RQ1: How effective is BERT fine-tuning for maritime terminology annotation when leveraging transfer learning from existing multi-domain annotated corpora?
- RQ2: Can GPT-4 effectively identify maritime safety terminology in a zero-shot setting through domain-specific prompts?

Experiment Results

Performance Comparison in Maritime Safety Domain

Metrics	BERT Fine-tuning			GPT-4 Zero-shot		
	Overall	B-tag	I-tag	Overall	B-tag	I-tag
Precision	32.8	29.2	65.5	49.1	57.6	30.5
Recall	15.6	19.9	8.3	36.9	40.3	27.4
F1-score	21.1	23.6	14.8	42.1	47.4	28.9

Key Findings

- **Experiment 1**: BERT Fine-tuning Results
  - Strong imbalance between precision (32.8) and recall (15.6) suggests challenges in cross-domain transfer
  - Higher I-tag precision (65.5) indicates potential in term continuation recognition
  - Limited overall effectiveness (F1: 21.1) highlights need for domain-specific adaptation
- **Experiment 2**: GPT-4 Zero-shot Results
  - Balanced precision-recall metrics (P: 49.1, R: 36.9) demonstrate stable domain adaptation
  - Strong B-tag performance (F1: 47.4) shows promise in term boundary identification
  - Overall F1-score (42.1) suggests potential of zero-shot learning in specialized terminology annotation

Future Research Directions

- Cross-lingual Terminology Extraction
  - Leverage multilingual pre-trained models for specialized domain terminology
  - Investigate cross-language term equivalence patterns
  - Study knowledge transfer across languages in domain-specific ATE
- Advanced LLMs for ATE Experiments
  - Compare LLMs' performance in recognizing domain-specific terminological patterns
  - Explore few-shot learning with term definition prompts
  - Investigate LLMs' ability in distinguishing term variations and relationships
  - Study contextual understanding of specialized terminology across domains
- Enhanced Annotation Frameworks
  - Evaluate alternative tagging schemes (IO vs IOB) for term boundary precision
  - Investigate nested term and split term annotation strategies
  - Research optimal annotation approaches for multi-word terms

Reference

Rigouts Terryn, A. (2021). ACTER terminology annotation guidelines. http://hdl.handle.net/1854/LU-8503113
Rigouts Terryn, A., Hoste, V., & Lefever, E. (2020). In no uncertain terms: A dataset for monolingual and multilingual automatic term extraction from comparable corpora. Language Resources & Evaluation, 54(2), 385–418. https://doi.org/10.1007/s10579-019-09453-9
Tran, H. T-H. (2024). Neural approaches to automatic terminology extraction [Doctoral dissertation, Jožef Stefan International Postgraduate School & La Rochelle University].

Training Data

- ACTER Annotated Corpora for Term Extraction Research (version 1.5) (Rigouts Terryn et al., 2020)
- Term categories: Specific Terms, Common Terms, Out-of-Domain Terms, Named Entities
- English corpus with IOB annotation scheme and Named Entities

Gold Standard Construction

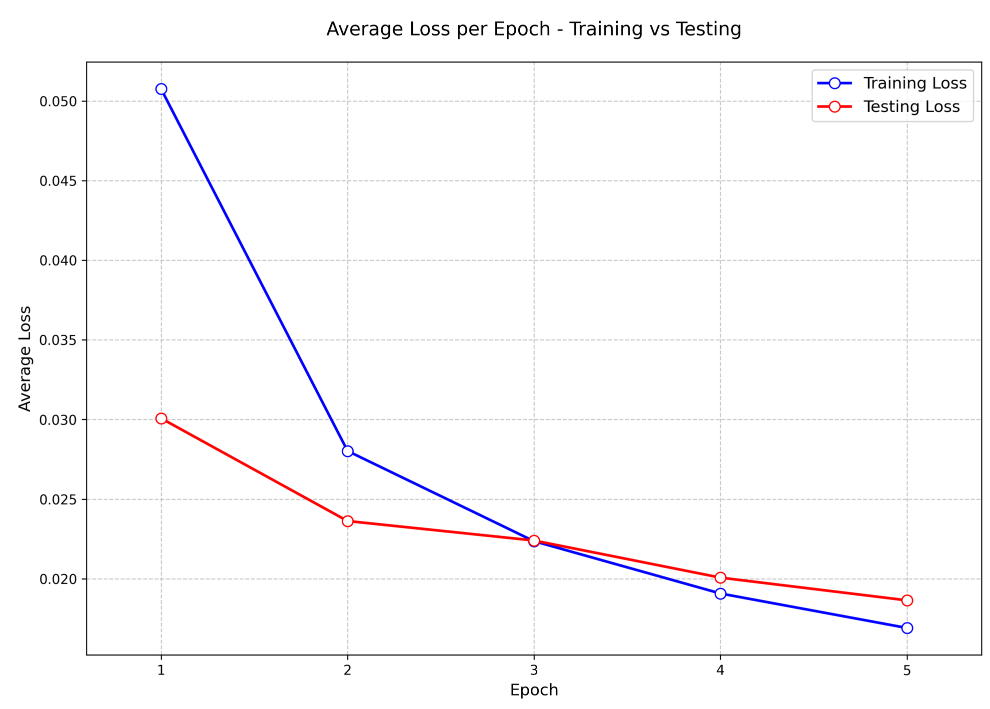
- Tool: Label Studio
- Text: Maritime Accident Report (tokenized) (9,737 tokens)
- Annotation Guidelines: Following ACTER terminology annotation guideline (Rigouts Terryn, 2021)

Term Type	Total Annotations	Unique Count	Duplication Rate	Percentage of Total Annotations
Common Terms	844	191	77.37%	47.68%
Specific Terms	462	145	68.61%	26.10%
Named Entities	340	80	76.47%	19.21%
Out-of-Domain Terms	124	34	72.58%	7.01%
Total	1,770	450	74.58%	100%

Experiment

Experiment 1: BERT Fine-tuning for Cross-domain Terminology Annotation

- Model Configuration
  - Base Model: BERT-base
  - Tokenizer: BertTokenizer
  - Maximum Sequence Length: 224
  - Batch Size: 16
  - Learning Rate: 5e-6
  - Number of Epochs: 5
  - Train-Test Split: 80%-20%



Metrics	Fine-tuned BERT Model Performance on ACTER Testing Set		
	Overall	B-tag	I-tag
Precision	67.4	71.2	49.4
Recall	70.5	83.1	35.0
F1-score	68.9	76.7	40.9

Experiment 2: GPT-4 Zero-shot Learning for Domain-specific Terminology Annotation

- Construct domain-specific prompt integrating ACTER's term classification framework and maritime expertise
- Configure GPT-4 API KEY with structured prompts to execute ATE

```
DOMAIN_DESCRIPTION = ```
You are an expert in maritime safety. Analyse the provided text for terms that are highly relevant to maritime security, considering both "lexicon-specificity" (whether only specialists understand the term) and "domain-specificity" (whether the term is primarily used in maritime contexts). Only classify terms that are crucial for understanding maritime safety and navigation. Classify each term into one of the following categories:
- Specific Terms: Terms that are both lexicon- and domain-specific; these terms are understood only by maritime specialists and are specific to the maritime field. Do not include general safety terms or concepts that are not unique to maritime contexts. Example: "COLREG."
- Out-of-Domain Terms (OOD): Technical terms that may be lexicon-specific (understood by specialists) but are not primarily related to maritime safety. Avoid categorising general industry or safety terms that are not unique to maritime operations. Example: "radar."
- Common Terms: Terms that are commonly understood but have a particular or specific meaning within maritime safety. These terms are part of general vocabulary but are commonly used in the maritime domain with specific implications. Avoid labelling terms that are generic and do not gain unique meaning in maritime contexts. Example: "vessel."
- Named Entities: Proper names of real-world objects, such as locations, organisations, or ship names that are directly related to the maritime domain. Avoid general names or entities unless they specifically relate to maritime safety. Example: "Port Adelaide."

Special Rules:
- Focus only on terms and entities that are directly relevant to maritime safety. Do not annotate general phrases or concepts that do not add significant meaning within the maritime context.
- Follow a recursive annotation approach: annotate key terms within compound phrases separately only if they hold specific meaning within the domain. Avoid over-annotating by splitting phrases unnecessarily.
- For descriptive phrases, only annotate adjectives if they convey a specific meaning within the maritime domain; otherwise, annotate only the main noun. For example, in "uncertified crew," annotate "crew" only.
- Ignore general terms with no specific maritime meaning, such as "incident" or "regulation," unless they are uniquely defined within the maritime context.
- Content inside parentheses or following slashes can be annotated separately if it qualifies as a term on its own, but do not combine it with the content outside. For instance, in "macro ( national / global ) levels," you may annotate "macro," "national," or "global" individually if relevant, but do not treat "macro ( national / global ) levels" as a single term.
```
```

GHENT
UNIVERSITY
```