Plausible-Parrots @ *MSP-WS2023*: Enhancing Semantic Plausibility Modeling Using Entity and Event Knowledge

Chong Shen, Chenyue Zhou

Institute for Natural Language Processing
University of Stuttgart
{*chong.shen, chenyue.zhou*}*@ims.uni-stuttgart.de*

Background and Motivation

- A crucial component for Natural Language Understanding: Discern plausible from implausible expressions
- Semantic Plausibility Modeling
  - fundamental element to understand event semantics (Resnik 1996)
  - identify events that are likely to happen but not necessarily attested (Gordon and Van Durme 2013)

Background and Motivation

Previous Work: Injecting world knowledge to ML models boosts the performance of semantic plausibility classification.

- Physical world knowledge (Wang, Durrett, and Erk 2018)
  - Sentience
  - Mass-count
  - Size
  - Weight
- Lexical hierarchy knowledge (Porada et al. 2021)
  - Hypernym chains of WordNet synsets
- Abstractness knowledge (Eichel and Im Walde 2023)
  - Integrate degrees of abstractness into event triple generation

Background and Motivation

- Events in real world scenarios: diverse, complex, dynamic form, e.g. natural language sentences.
- Ambiguity: ”Jobs takes an apple.”
  - ”Jobs”
  - ”apple”
  - ”take”
- Types of entities and events (verbs) matter!
- Knowledge injection via template:
  Natural language sentences as inputs fit better with LLMs.
- **Our work:** Inject **entity type** and **event type** as external knowledge

Related Work

Ultra Fine-grained Entity Typing (UFET)

- Multi-label classification problem
- Predict fine-grained semantic types for entity mentions in text (Choi et al. 2018)

Sentence with Target Entity	Entity Types
During the Inca Empire, {**the Inti Raymi**} was the most important of four ceremonies celebrated in Cusco.	event, festival, **ritual, custom, ceremony, party, celebration**
{**They**} have been asked to appear in court to face the charge.	person, **accused, suspect, defendant**
Ban praised Rwanda's commitment to the UN and its role in {**peacemaking operations**}.	event, **plan, mission, action**

Examples of entity mentions and their annotated types.

Related Work

Simple vs. Complex Event Representation

- Simple event representation: (Balasubramanian, Soderland, Etzioni, et al. 2013)
  (subject, verb, object)-triples
- Rich event representation: (Li et al. 2022)
  (event trigger, event type, event argument, argument role)

Event Detection (ED)

- Identifying event triggers and classifying them into event types (Li et al. 2023)

Task Definitions

- **Research Question:** Do event type- and entity type knowledge help improve semantic plausibility classification?
- **Assumption:** Entity types, event types and their definitions provide rich semantic knowledge and boost the model's performance on predicting plausibility.
- **Task Formulation**: Binary sequence classification problem
  - Input: a knowledge-enhanced event sentence as prompt
  - Output: *plausible (1)* or *implausible (0)*

Methods

1. **UFET**
   - Input: subject *s* and object *o* of an *(s,v,o)*-event triple
   - Output: a list L_s and L_o of fine-grained types for each entity (i.e. *s* and *o*)
   - Model: **CASENT** (Feng, Pratapa, and Mortensen 2023)
     - A Seq2Seq model with calibrated confidence for UFET
2. **ED**
   - Input: a natural language sentence derived from an *(s,v,o)*-triple
   - Output: a type label t_v for the event trigger *v*
   - Model: **CEDAR** (Li et al. 2023)
     - A general purpose event detection model for thousands of types

Methods

③ Template-based Prompt Engineering

- Input:
  an event sentence $concat(Svo.)$ +
  entity types $L_s = \{t_{s1}, t_{s2}, ...\}$ of s +
  entity types $L_o = \{t_{o1}, t_{o2}, ...\}$ of o +
  event type t_v +
  definitions for t_{si}, t_{oi}, and t_v extracted from a KB [1]
- Output: a knowledge-enhanced event sentence as prompt p

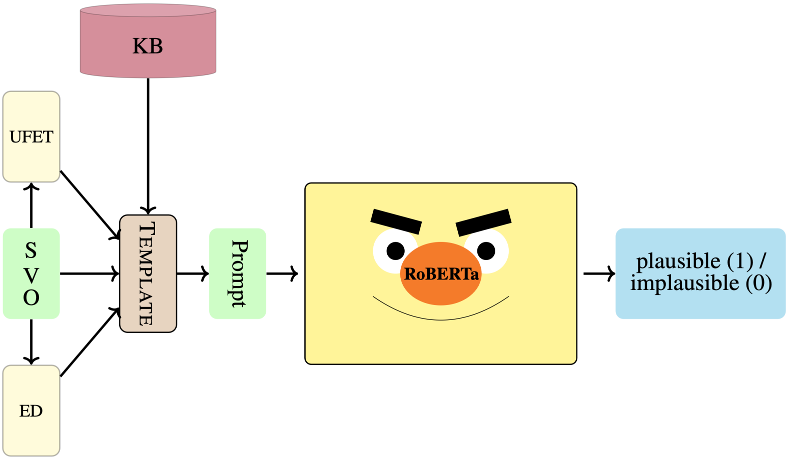
[1] We use WikiData: https://www.wikidata.org/

Methods

3 Template-based Prompt Engineering

- Templates: See Appendix. Go to Appendix
- Example prompt for the event *(county, receives, hour)*:
  [EVT]County receives hour.[/EVT]
  The subject "County" has type [STYPE]municipality[/STYPE], which means [DEF]administrative division having corporate status and usually powers of self-government or jurisdiction[/DEF]. It can also have type [STYPE]place[/STYPE], which means ...
  The verb "receives" has type [ETYPE]reception[/ETYPE], which means [DEF]process of recovering a message, signal, or item that has been transmitted[/DEF].
  The object "hour" has type [OTYPE]hour[/OTYPE], which means [DEF]unit of time[/DEF]. It can also have type [OTYPE]time[/OTYPE], which means ...

System Overview



Experiments

Datasets

Preliminary Study

- Statistic analysis
  - Number of examples
  - Number of each feature (given by column name)
  - Average/minimal/maximal length, variance, standard deviation of textual features
  - Frequency and distribution of features (e.g., modifiers, nouns and labels in Adept)

Experiments
Datasets

Preliminary Study

- Linguistic analysis
  - Stop words
  - N-grams
  - POS tags
  - Wordcloud that illustrates the most frequent words associated with each label
  - Semantic similarities between these words. This could help identify whether certain types of verbs or objects are more associated with plausible or implausible labels.

Experiments

Datasets

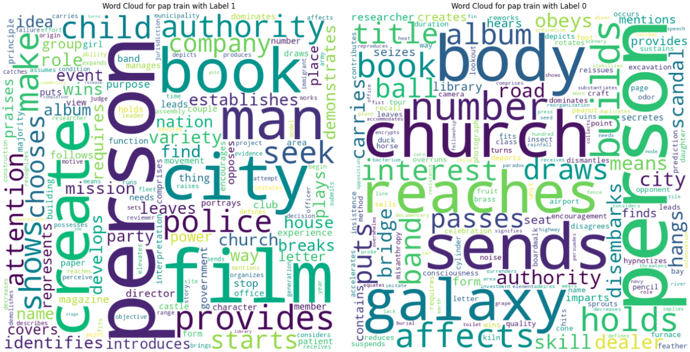


Figure: Wordcloud of PAP train (Left: Plausible Right: Implausible).

Experiments

Datasets

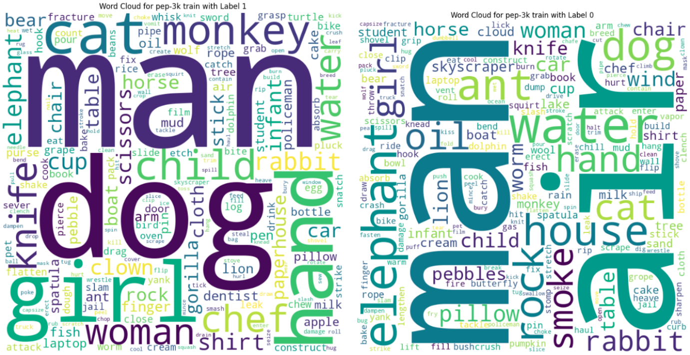


Figure: Wordcloud of PEP-3K train (Left: Plausible Right: Implausible).

Experiments
Datasets

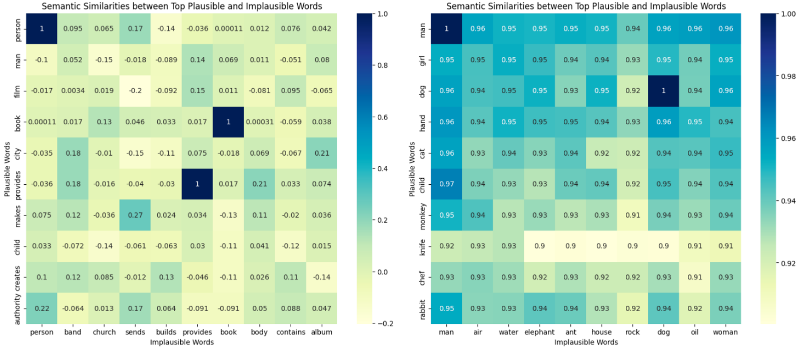


Figure: Word similarity between top plausible words and implausible words. (Left: PAP Right:PEP-3K)

Experiments

Datasets

Data Augmentation

- Label imbalance problem
- Randomly took data points labelles as 'implausible' from PAP multiclass dataset, added 734 datapoints to PAP train, 92 to PAP dev, 92 to PAP test
- PEP datasets are unchanged since they are more evenly distributed

Experiments

Datasets

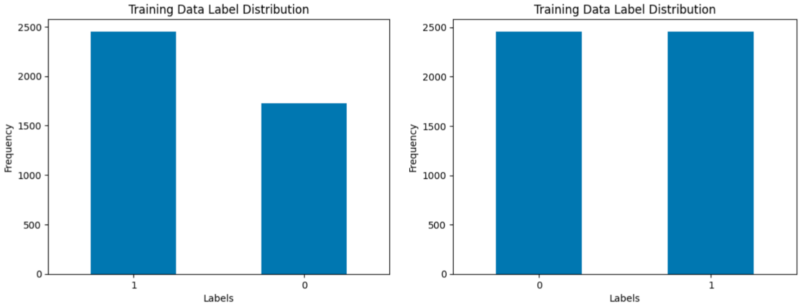


Figure: Traingin Data Label Distribution - Before and After Data Augmentation.

Experiments

Baselines

1. **Weak Baseline:** Zero-shot Inference
   - Model: RoBERTa-large [2]
   - Data: event sentences without additional knowledge
2. **Strong Baseline:** Fine-tuning
   - Model: RoBERTa-large
   - Data: event sentences without additional knowledge

[2] https://huggingface.co/roberta-large

Experiments

Settings

- Model: RoBERTa-large
- Model parameters: 355 Million
- Optimizer: AdamW
- Training hyperparameters:
  - number of epochs: 10
  - batch size: 16
  - learning rate: 1e-5
  - warmup steps: 0.01
  - weight decay: 10

Results

Model	Pap					Pep				
	AUC	P	R	F1	Acc	AUC	P	R	F1	Acc
RoBERTa$_{evt+ent}$	0.659	0.717	0.526	0.607	0.659	**0.883**	**0.850**	**0.928**	**0.888**	**0.883**
RoBERTa$_{evt}$	0.636	0.640	**0.623**	**0.632**	0.636	0.844	0.793	**0.928**	0.855	0.844
RoBERTa$_{ent}$	**0.666**	**0.763**	0.481	0.590	**0.666**	0.840	0.838	0.843	0.840	0.840
RoBERTa$_{bs,0-shot}$	0.532	0.564	0.286	0.379	0.532	0.500	0.500	0.137	0.215	0.502
RoBERTa$_{bs,ft}$	0.646	0.737	0.455	0.562	0.646	0.791	0.830	0.732	0.778	0.792

- For **Pap**, injecting **entity type** leads to the best AUC (0.666).
- For **Pep**, injecting **both** entity type and event type significantly improves all metrics.

Results

Model	Pap					Pep				
	AUC	P	R	F1	Acc	AUC	P	R	F1	Acc
RoBERTa$_{evt+ent}$	0.659	0.717	0.526	0.607	0.659	**0.883**	**0.850**	**0.928**	**0.888**	**0.883**
RoBERTa$_{evt}$	0.636	0.640	**0.623**	**0.632**	0.636	0.844	0.793	**0.928**	0.855	0.844
RoBERTa$_{ent}$	**0.666**	**0.763**	0.481	0.590	**0.666**	0.840	0.838	0.843	0.840	0.840
RoBERTa$_{bs,0-shot}$	0.532	0.564	0.286	0.379	0.532	0.500	0.500	0.137	0.215	0.502
RoBERTa$_{bs,ft}$	0.646	0.737	0.455	0.562	0.646	0.791	0.830	0.732	0.778	0.792

- For both datasets:
  - Finetuning without additional knowledge improves all metrics compared to zero-shot inference
  - Injecting event type improves the recall but decreases the precision.
  - Injecting entity type improves the precision but decreases the recall.

Qualitative Analysis

Pap (evt+ent, fine-tune):

- 105 wrong predictions among 308 test examples
- 68 examples with unknown verb type

```
y_pred: 0 y_true 1
[EVT]Trader ensures strategy.[/EVT]
The subject "Trader" has type [STYPE]person[/STYPE], which means [DEF]being that has certain capacities or attribut
es constituting personhood (avoid use with P31; use Q5 for humans)[/DEF]. It can also have type [STYPE]businessperso
on[/STYPE], which means [DEF]person involved in activities for the purpose of generating revenue[/DEF]. It can also
have type [STYPE]trader[/STYPE], which means [DEF]businessperson who exchanges stocks, bonds and other such financi
al instruments[/DEF]. It can also have type [STYPE]professional[/STYPE], which means [DEF]person who is paid to und
ertake a specialized set of tasks and to complete them for a fee[/DEF]. It can also have type [STYPE]entrepreneur[/
STYPE], which means [DEF]individual who organizes and operates a business[/DEF].
The verb "ensures" has an unknown type.
The object "strategy" has type [OTYPE]idea[/OTYPE], which means [DEF]mental image or concept[/DEF]. It can also hav
e type [OTYPE]concept[/OTYPE], which means [DEF]semantic unit understood in different ways, e.g. as mental represen
tation, ability or abstract object[/DEF]. It can also have type [OTYPE]plan[/OTYPE], which means [DEF]outline of a
strategy for achievement of an objective[/DEF].
```

Qualitative Analysis

Pap (evt+ent, fine-tune):

- 105 wrong predictions among 308 test examples
- 68 examples with unknown verb type
- 50 examples with trivial entity type ”entity”

```
y_pred: 1 y_true 0
[EVT]Hook wins role.[/EVT]
The subject "Hook" has type [STYPE]concept[/STYPE], which means [DEF]semantic unit understood in different ways, e.
g. as mental representation, ability or abstract object[/DEF]. It can also have type [STYPE]idea[/STYPE], which mea
ns [DEF]mental image or concept[/DEF]. It can also have type [STYPE]entity[/STYPE], which means [DEF]anything that
can be considered, discussed, or observed[/DEF]. It can also have type [STYPE]hook[/STYPE], which means [DEF]object
for hanging, fishing etc.[/DEF].
The verb "wins" has an unknown type.
The object "role" has type [OTYPE]entity[/OTYPE], which means [DEF]anything that can be considered, discussed, or o
bserved[/DEF]. It can also have type [OTYPE]role[/OTYPE], which means [DEF]set of behaviours, rights, obligations,
beliefs, and norms expected from an individual that has a certain social status[/DEF].
```

Conclusions

1. We enhance a large language model with fine-grained entity types, event types and their definitions as external knowledge.
2. We carefully design templates to integrate these external knowledge into the input.
3. We deal with unbalanced label distribution via data augmentation.
4. We adapt the task to real world scenarios in which the inputs are natural language sentences (rather than (s,v,o)-triples).
5. Injecting entity and event type knowledge indeed benefits model's performance on the task.
6. The results on Pap do not fully meet our expectation.

Future Work

1. Other knowledge (e.g., semantic roles)
2. Other modalities (e.g., image, video)

Thank you!

References I

- Balasubramanian, Niranjan, Stephen Soderland, Oren Etzioni, et al. (2013). "Generating coherent event schemas at scale". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1721–1731.
- Choi, Eunsol et al. (2018). "Ultra-fine entity typing". In: *arXiv preprint arXiv:1807.04905*.
- Eichel, Annerose and Sabine Schulte Im Walde (2023). "A Dataset for Physical and Abstract Plausibility and Sources of Human Disagreement". In: *Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*, pp. 31–45.
- Feng, Yanlin, Adithya Pratapa, and David R Mortensen (2023). "Calibrated Seq2seq Models for Efficient and Generalizable Ultra-fine Entity Typing". In: *arXiv preprint arXiv:2311.00835*.
- Gordon, Jonathan and Benjamin Van Durme (2013). "Reporting bias and knowledge acquisition". In: *Proceedings of the 2013 workshop on Automated knowledge base construction*, pp. 25–30.

References II

- Li, Qian et al. (2022). "A survey on deep learning event extraction: Approaches and applications". In: *IEEE Transactions on Neural Networks and Learning Systems*.
- Li, Sha et al. (2023). "GLEN: General-Purpose Event Detection for Thousands of Types". In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 2823–2838.
- Porada, Ian et al. (2021). "Modeling event plausibility with consistent conceptual abstraction". In: *arXiv e-prints*, arXiv–2104.
- Resnik, Philip (1996). "Selectional constraints: An information-theoretic model and its computational realization". In: *Cognition* 61.1-2, pp. 127–159.
- Wang, Su, Greg Durrett, and Katrin Erk (2018). "Modeling semantic plausibility by injecting world knowledge". In: *arXiv preprint arXiv:1804.00619*.

Appendix - Template and Example Prompt

Back to methods

1. TEMPLATE_SENT = [EVT]{sent}[/EVT]
2. TEMPLATE_SUBJ_BASIC = The subject "{subj}" has type [STYPE]{stype}[/STYPE], which means [DEF]{stype_desc}[/DEF].
3. TEMPLATE_SUBJ_EXTEND = It can also have type [STYPE]{stype}[/STYPE], which means [DEF]stype_desc[/DEF].
4. TEMPLATE_SUBJ_UNK = The subject "{subj}" has an unknown type.
5. TEMPLATE_VERB = The verb "{verb}" has type [ETYPE]etype[/ETYPE], which means [DEF]etype_desc[/DEF].
6. TEMPLATE_VERB_UNK = The verb "{verb}" has an unknown type.
7. TEMPLATE_OBJ_BASIC = The object "{obj}" has type [OTYPE]otype[/OTYPE], which means [DEF]otype_desc[/DEF].
8. TEMPLATE_OBJ_EXTEND = It can also have type [OTYPE]otype[/OTYPE], which means [DEF]otype_desc[/DEF].
9. TEMPLATE_OBJ_UNK = The object "{obj}" has an unknown type.