Paper ID: 4195

EMORED: A DATASET FOR RELATION EXTRACTION IN TEXTS WITH EMOTICONS

Lingxing Kong[1], *Zheng Ma*[1], *Jianbing Zhang*[1,2], *Liang He*[1], *Jiajun Chen*[1]

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

[2]School of Artificial Intelligence, Nanjing University, China

Sentence in NYT:

Mason Adams was born in Brooklyn

Relational triple:

Adams, place_of_birth, Brooklyn

Sentence with emoticons in the Real World:

Eric Drooker is an American painter, graphic novelist, and frequent cover artist for The New Yorker.

Relational triples:

Eric Drooker, occupation, painter

Eric Drooker, occupation, novelist

Eric Drooker, occupation, artist

Motivation

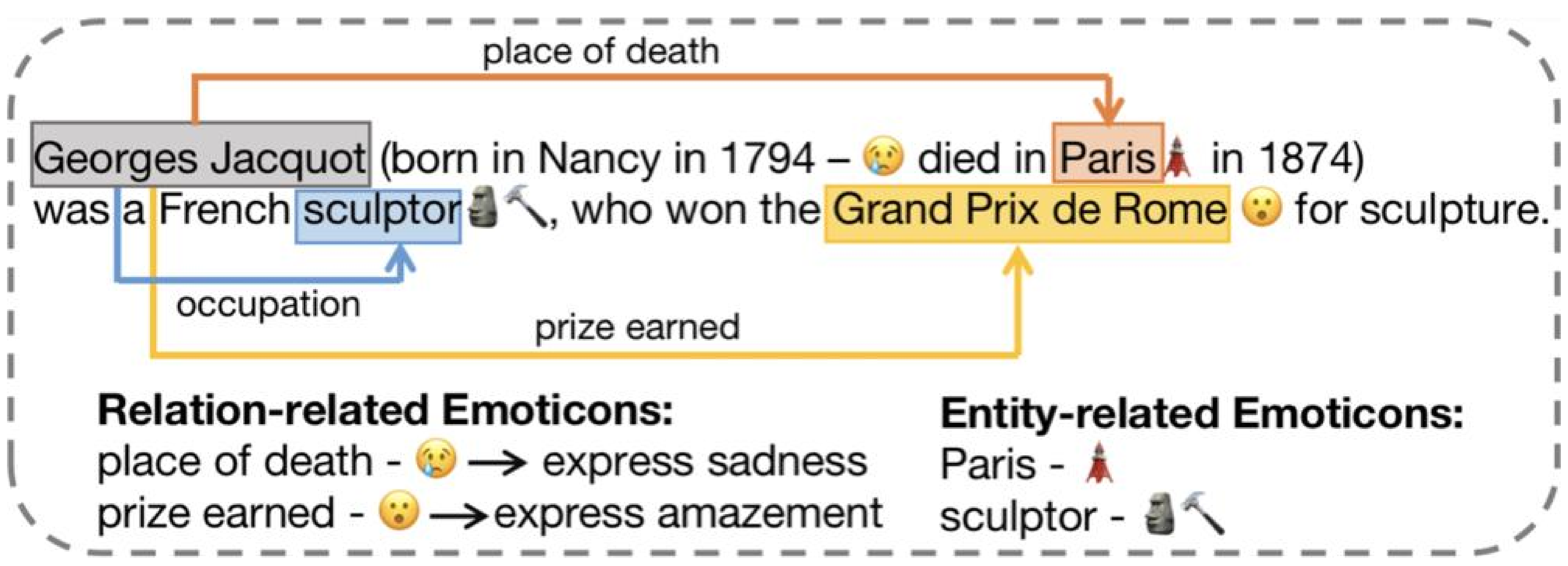
- Sentences in real-world scenarios tend to be more complex, containing multiple relational triples and various emoticons, in stark contrast to the simplistic sentences found in existing RE datasets.
- Are SOTA RE models robust in emoticon-rich contexts?
- Can the RE models harness the information encoded within emoticons to enhance their performance ?

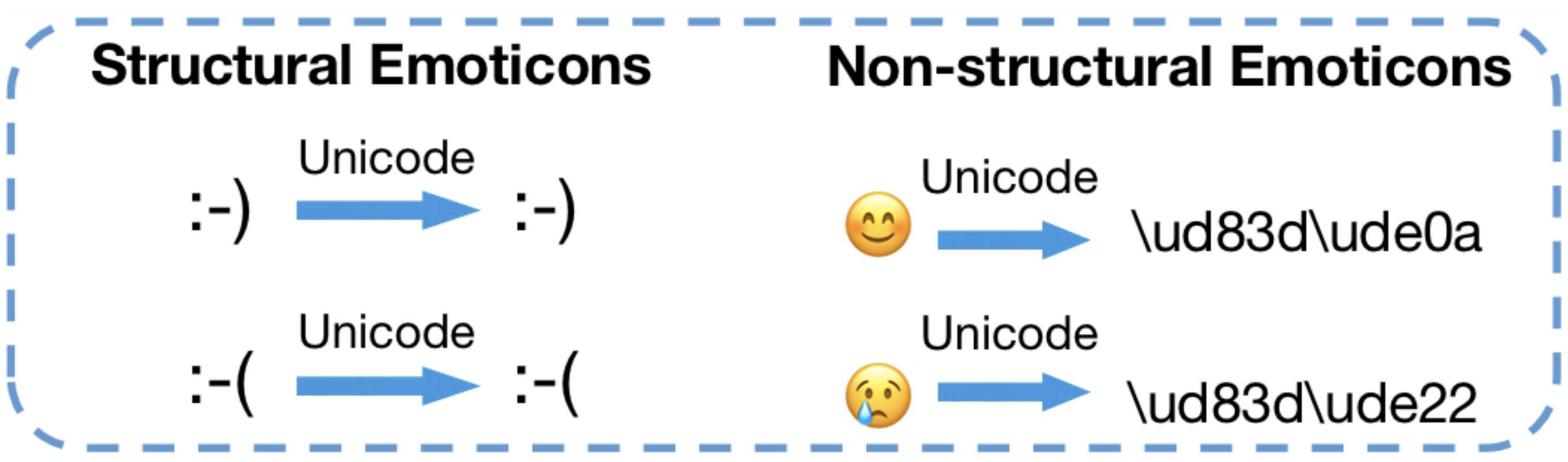
Challenge

- Lacking the dedicated dataset for the Emoticon-infused Relation Extraction (EmoRE) task.
- The emoticon-infused samples should be suitable for evaluating SOTA relation extraction models.
- Considering different emoticon types and their impacts on model performance for the EmoRE task.

Dataset Construction

- When collecting the sample candidates, we use Wikipedia as our primary corpus, with Wikidata serving as our knowledge base.
- We manually add emoticons into the sample candidates based on the functional roles of the emoticons.
- We further categorize emoticons into two types: structural and non-structural, based on their behavior under the Unicode mode.





Dataset Statistics

Datasets	#Text	#Rel.	#Triple	Emoticon Types	#Ent. Per Sent.	#Rel. Per Sent.
SemEval	13,434	10	13,434	-	2.0	1.0
NYT	66,194	24	104,339	-	1.0	0.8
WebNLG	6,222	171	14,485	-	1.3	0.9
TACRED	106,264	41	21,773	-	2.0	1.0
EmoRED	13,856	40	29,374	**2**	**3.1**	**2.1**

Experiments

- We evaluate both state-of-the-art supervised models and Large Language Models (LLMs) on EmoRED.
- We investigate the impact of different emoticon types on model performance.
- For LLMs, we compare seven popular models, shedding light on their performance and explanatory capabilities within EmoRE.
- We identify novel exemplars for enhancing LLM performance.

	NYT	EmoRED-none	EmoRED
Joint Models	**F1**	**F1**	**F1**
NovelTagging	42.0	37.8	42.8
CasRel	89.6	76.1	76.7
	TACRED	**EmoRED-none**	**EmoRED**
Pipeline Models	**F1**	**F1**	**F1**
BERT	66.4	71.5	70.5
SpanBERT	70.8	73.3	72.2
PURE	-	73.9	72.7
Large Language Models	**F1**	**F1**	**F1**
GPT-4	-	26.9	27.4
GPT-3.5	20.3	14.3	14.2
ChatGLM	-	4.3	5.9
BLOOMZ	-	1.0	1.2
Llama-2-7b	-	0.5	0.5
Llama-2-13b	-	3.1	2.6
XVERSE	-	5.6	5.5

	Structual	Non-Structual
Models	**F1**	**F1**
NovelTagging	27.0	36.7
CasRel	73.4	75.0
PURE	69.9	74.0
BERT	68.3	70.8
SpanBERT	69.6	73.3
GPT-4	26.7	28.1
GPT-3.5	13.8	14.5
ChatGLM	3.9	6.5
BLOOMZ	0.9	1.4
Llama-2-7b	0.6	0.4
Llama-2-13b	2.0	3.0
XVERSE	5.6	5.5

Results and Analysis

- Among supervised models, joint models derive benefits from the illustrative information provided by emoticons while all pipeline models fail to do.
- Supervised models consistently outperform LLMs in both the emoticon and non-emoticon settings.
- Most evaluated models perform better in the non-structural type than in the structural type.
- Exemplars featuring relation-related emoticons play a significant role in enhancing the performance of LLMs.