

Fine-grained Control of Affective Natural Language Generation through Appraisal Conditions

Yarik Menchaca Resendiz and Roman Klinger
{Yarik.Menchaca-Resendiz,Roman.Klinger}@ims.uni-stuttgart.de

1. Motivation

Generate emotional connotated textual (implicit or explicit) event description.

"My dog ran away, and we couldn't find him for hours."

Limitation of models conditioned on basic emotions or valence and arousal values:

- They struggle to communicate emotions implicitly.

Advantages of including appraisal theories in addition to emotion theories:

- Explicitly link events and their evaluation to emotions.
- Fine-grained control over the generated text.

2. Methods

Model configurations built on top of Bart and T5.

- E: (condition on emotions only)

"generate [emotion]: [prompt]"

- EA: (condition on both emotions and appraisals)

"generate [emotion] [appraisals]: [prompt]"

- A: (condition on appraisals only)

"generate [appraisal]: [prompt]"

Conf	Input Example	Output Example
E	generate joy : <i>Las day I</i>	was very relaxed.
EA	generate joy attention NoRESP control NoCIRC NoPLEA effort NoCERT : <i>Last day I</i>	was very relaxed because I worked for 6 hours
A	generate attention NoRESP control NoCIRC NoPLEA effort NoCERT : <i>Last day I</i>	decided to work for 6 hours

Table 1. Input and output representation for the emotion Joy, and the appraisals Attention, Control and Effort. Conditions are in bold (emotions and appraisals) and the prompt is in italic.

3. Experiments

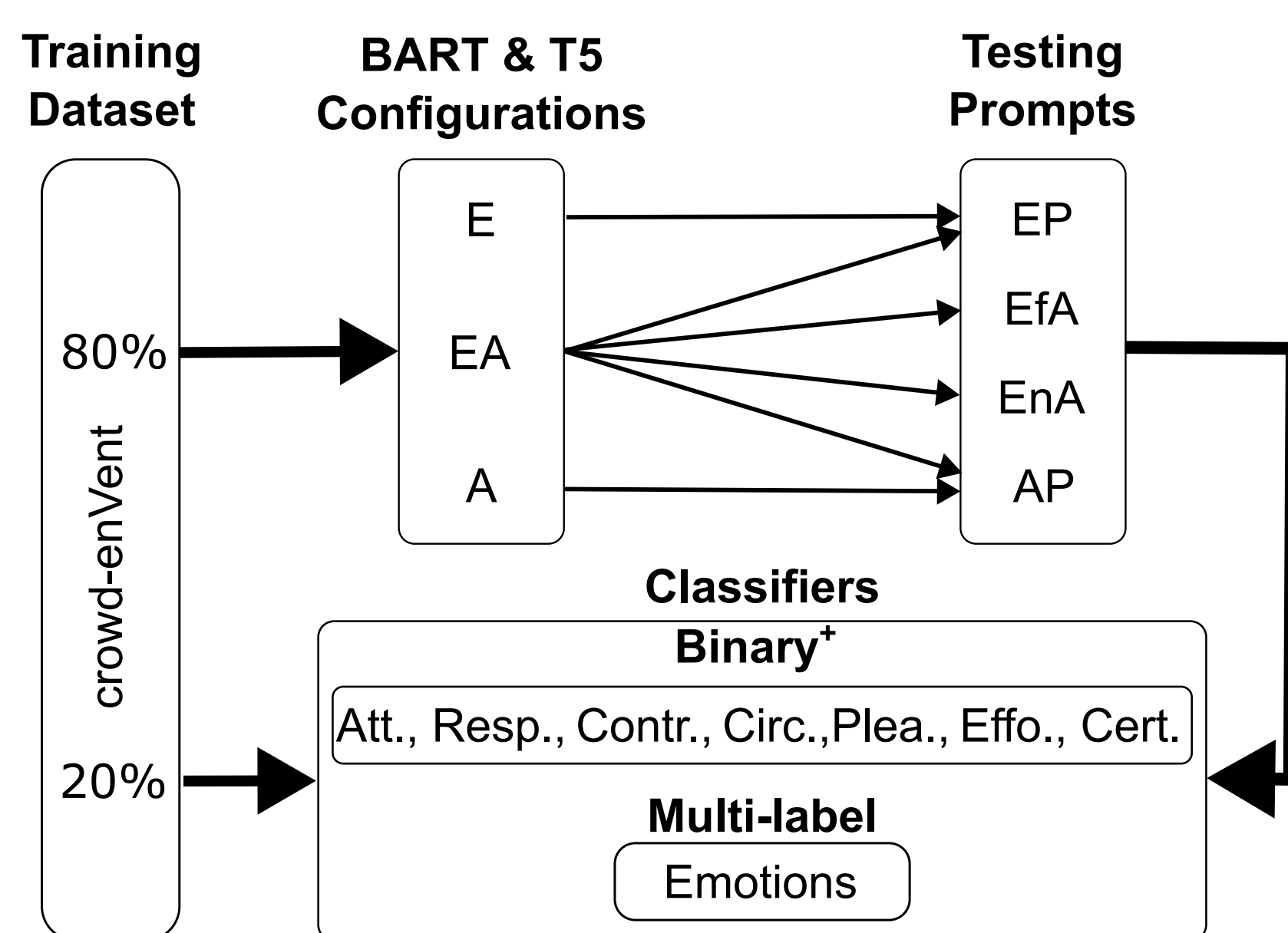


Figure 1. Experiment workflow for the crowd-enVent dataset, CNLG models, classifiers, and evaluation datasets.

4. Results

RQ1: Do Appraisal Variables Improve Affective Text Generation?

Arch.	Conf.	Testing Prmpt.	Anger	Disg.	Fear	Guilt	Joy	Sad.	Shame	M. Avg.
ATG.	E		0.10	0.18	0.25	0.06	0.17			0.15
T5	E	EP	0.28	0.50	0.63	0.23	0.60	0.32	0.40	0.42
T5	EA	EP	0.46	0.58	0.70	0.27	0.77	0.58	0.32	0.52
T5	EA	EfA	0.39	0.60	0.57	0.35	0.77	0.47	0.21	0.48
T5	EA	EnAP	0.52	0.55	0.64	0.35	0.58	0.41	0.19	0.46
Bart	E	EP	0.36	0.45	0.40	0.29	0.63	0.43	0.49	0.43
Bart	EA	EP	0.41	0.57	0.48	0.41	0.63	0.54	0.36	0.49
Bart	EA	EfA	0.34	0.45	0.52	0.29	0.75	0.46	0.44	0.47
Bart	EA	EnAP	0.34	0.51	0.43	0.26	0.57	0.33	0.37	0.40

Table 2. Emotion F1 scores of models trained with only emotions (E), emotions and appraisal conditions (EA), and only appraisal conditions (A) over the generated text using the testing prompt sets: EP, EnAP, and EfA.

RQ2: Do Appraisal Variables Improve Affective Text Generation?

N.	Conf.	Condition	Prompt	Generated Text
1	E	Joy	I got	I got a job I really wanted.
2	EA	Joy, Resp., Contr., Plea.	I got	I got a job I had wanted for months leading up to my graduation.
3	EA	Joy, Resp., Contr., Plea. Effo.	I got	I got a job I applied for last week.
4	E	Sadness	When my	When my dad passed away
5	EA	Sadness, Effo., Cert.	When my	when my boyfriend told me he was leaving for college because I was too busy to do anything with him
6	EA	Sadness, Plea.	When my	when my boyfriend told me he was leaving for college because I felt alone and I like I was going too
7	E	Anger	I found	I found out my partner was cheating on me
8	EA	Anger, Plea.	I found	I found out my partner was cheating on me
9	EA	Anger, Effo., Cert.	I found	I found out my partner was taking advantage of my offer to buy him a house.

Table 3: Example texts generated by T5 using different model configurations, conditions, and Trigger-Phrases.

5. Human Evaluation

	Conf.	Testing Prmpt.	Anger	Disgust	Fear	Guilt	Joy	Sadness	Shame	M. Avg.
Hum.	Hum.	enVent	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E	EP	0.69	0.72	0.72	0.83	0.89	0.67	0.82	0.76
	EA	EP	0.79	0.74	0.73	0.62	0.92	0.82	0.60	0.74
Auto.	EA	EfA	0.73	0.67	0.62	0.45	0.71	0.74	0.65	0.65
	Hum.	enVent	0.86	1.00	0.90	1.00	1.00	1.00	1.00	0.97
	E	EP	0.46	0.14	0.50	0.44	0.78	0.33	0.41	0.44
EA	EP	0.55	0.38	0.82	0.31	1.00	0.60	0.26	0.56	
	EfA	0.53	0.50	0.33	0.40	0.67	0.50	0.20	0.45	

Table 4. Human annotation results as F1 (top). For comparison, we show the automatic evaluation on the same subsample (bottom).

Conf.	Testing Prmpt.	Fluency	Grammar	Native Spkr.	Coherency	Really Happen	Written by AI	Written by Human
Hum.	enVent	4.10	2.98	4.00	3.83	4.47	2.83	3.92
E	EP	3.55	2.43	3.40	3.36	4.00	2.42	3.25
EA	EP	3.07	1.88	2.82	2.89	3.57	1.86	2.93
EA	EfA	3.55	2.43	3.30	3.23	3.88	2.17	3.18

Table 5. Human evaluation of text quality using the five-level Likert scale, where 1 is "not agree at all", and 5 is "extremely agree". (higher is better).

6. Conclusion and Future Work

- First study on conditional text generation based on more than one emotion model (basic emotions and appraisal).
- Adding appraisal information results in a 10pp improvement in text generation for a target emotion.
- Appraisal conditions encourage CNLG models to add details to event descriptions.
- *Additional conditions (e.g., topics or a previous utterance in a dialogue).*
- *Compatibility between conditions (emotions and appraisals).*

Acknowledgements

This work has been supported by a CONACYT scholarship (2020-000009-01EXTF-00195) and by the German Research Council (DFG), project "Computational Event Analysis based on Appraisal Theories for Emotion Analysis" (CEAT, project number KL 2869/1-2).